JOURNAL OF APPLIED PHYSICS

JOURNAL OF APPLIED PHYSICS

Vol. 21, No. 10 OCTOBER, 1950

In This Issue

		Page
Effect of Input Configuration on Antenna Impedance	J. R. Whinnery	945
Microwave Techniques for the Measurement of the Dielectric Constant of Fibers and Films of High Polymers	T. M. Shaw and J. J. Windle	956
Otto A. Beeck		961
Determination of Residual Stresses of Quenching Origin in Solid and Concentric Hollow Cylinders from Interferometric Observations	A. W. Sáenz	962
Some Crossover Properties in the Electron Immersion Objective	L. Jacob	966
Single Crystal Copper Surfaces	T. N. Rhodin, Jr.	971
Theory of the Parallel Plane Diode	A. H. Taub and Nelson Wax	974
High Frequency Impedance of Low Pressure Gaseous Diodes	Chai Yeh and E. L. Chaffee	981
Propagation of Plastic Deformation in Solids	Theodore von Karman and Pol Duwez	987
Kerr Cell Camera and Flash Illumination Unit for Ballistic Photography	H. F. Quinn, W. B. McKay, and O. J. Bourque	995
Effect of a Circular Groundplane on Antenna Radiation	A. Leitner and R. D. Spence	1001
Turbulence in Apparatus for Measurement of Streaming Double Refraction	H. G. Jerrard	1007
Effective Stress and Effective Strain in Relation to Stress Theories of Plasticity	G. N. White, Jr. and D. C. Drucker	1013
Formation of Crystal Nuclei in Liquid Metals	D. Turnbull	1022
Hot-Cathode Arcs in Cesium Vapor	Richard K. Steinberg	1028
Allotropy of Beryllium	S. S. Sidhu and C. O. Henry	1036
Single Wave-Length X-Rays for Powder Diffraction	L. K. Frevel and P. P. North	1038
Shielding Properties of the Concrete Wall of the M. I. T. Cyclotron	Victor Delano and Clark Goodman	1040
Storage of Small Signals on a Dielectric Surface	John V. Harrington	1048
Graphical Representation of Particle Trajectories in a Moving Reference System	M. Garbuny	1054
Some Observations on the Epitaxy of Sodium Chloride on Silver	G. W. Johnson	1057
Letters to the Editor:		
Note on Stability of Electron Flow in the Presence of Positive Ions	J. R. Pierce	1063
Modification of the Radiofrequency Mass Spectrometer	A. F. Henson	1063
Remark on an Aberration-Phenomenon in Electrostatic Lenses, "Kidney Effect"	E. Bas-Taymaz	1064
Dielectric Constants of Five Gases at 9400 Megacycles per Second	James A. Murphy and Richard C. Raymond	1064
Simplified Air Flow Method for the Determination of the Surface Area of Powders	Gerard Kraus and John R. Thiem	1065
Pursuit Course	Ralph Hoyt Bacon	1065
Some Observations on the Back Heating of Magnetron Cathodes	R. T. Young, Jr., L. W. Holmboe, and W. E. Waters, Jr.	1066
Some Magneto-Acoustic Effects in Nickel	T. F. Rogers and S. J. Johnson	1067
Electron-Microradiography of Electrodeposited Metals	Rolf Weil and Harold J. Read	1068
Limiting Hydrostatic Tension of Water Near 0°C	J. C. Fisher	1068
Local Order in Solid Cu_3Au	R. A. Oriani	1068
Formation of Annealing Twins during Grain Growth	R. L. Fullman	1069

BOARD OF EDITORS

Elmer Hutchisson, *Editor*

Earl C. Gregg, *Assistant Editor*

Associate Editors

Howard H. Aiken
James Hillier
W. James Lyons
Philip Morrison
Eric Reissner
Raymond J. Seeger
Frederick Seitz
Charles S. Smith

Publication of the

American Institute of Physics

OFFICERS

George R. Harrison, *Chairman*
George B. Pegram, *Treasurer*

GOVERNING BOARD

J. W. Beams
C. Paul Boner
Wallace R. Brode
W. F. Fair, Jr.
J. B. Fisk
G. P. Harnwell
George R. Harrison
W. V. Houston
Rudolph Kingslake
Paul Kirkpatrick
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
J. R. Oppenheimer
George B. Pegram
Harold K. Schilling
Francis G. Slack
John C. Steinberg

Consultant on Printing

Melvin Loos

ADMINISTRATIVE STAFF

Henry A. Barton, *Director*
Wallace Waterfall, *Executive Secretary*
Theodore Vorburger, *Assistant Treasurer and Comptroller*
Ruth F. Bryans, *Publication Manager*
Betty H. Goodfriend, *Assistant Secretary*
Eileen B. Neuberger, *Circulation Manager*

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $10.00. Single copies: $1.00 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 21, Number 10 October, 1950

The Effect of Input Configuration on Antenna Impedance*

J. R. Whinnery
University of California, Berkeley, California
(Received February 13, 1950)

It is known that the configuration of the input driving system may be of importance to the impedance of antennas at high frequencies. In this paper, a region in the vicinity of the input is considered as a transducer between the TEM mode on the driving transmission line and the spherical nearly-TEM mode on the antenna system, and the constants of this transducer are found both by calculation and measurement for a particular case. The calculation is based upon quasi-static concepts, utilizing an electrostatic field map obtained from an electrolytic tank to give the approximate form of the electric field lines. The measurement is made by finding three input reactances corresponding to the input region closed by three spun aluminum hemispheres of different radii. The three sets of data, taken at each frequency, are sufficient to determine the constants of the transducer. They are expressed here as a pi-network for comparison with calculated values.

The main portion of the antenna is analyzed substantially as in Schelkunoff's approach, with the TM modes in space and in the antenna region expressed as a shunt admittance at the end of the antenna in the TEM mode equivalent circuit. This admittance is then transformed to the desired reference on the antenna by a perturbation calculation. From there it is transformed to a desired reference on the feed line by means of the input network constants. Values calculated in this way are compared with measured values over the range 200 to 1000 mc/sec. for the particular monopoles studied. Agreement is not perfect, but appears considerably better than without consideration of the specific input configuration.

INTRODUCTION

IT is true that in many practical antenna problems, especially at frequencies below 100 mc/sec., the actual input driving means may add only a small contribution to input admittance. In these cases, it is not necessary to consider the actual input configuration in a mathematical analysis of the antenna, but, if an infinitesimal driving gap is assumed, the infinite susceptance singularity which would ordinarily result from this assumption must be avoided in some way. It will be described in the following paragraphs how this is accomplished in three of the modern methods of antenna analysis.

For frequencies above 100 mc/sec. (and for some antennas below this range), the actual input configuration may be important to the input impedance of an antenna. If the local capacitance of the actual input system is known, a first approximation may be had by adding the corresponding susceptance to the input admittance calculated by any of the analyses which have removed the local effects by avoiding the singularity. As an example, in the measured curves of Brown and Woodward,[1] their Fig. 13 shows curves of input resistance *versus* antenna length for antennas identical except with respect to their input configurations. These curves differ widely. They were able to explain much of the difference in these results in terms of an effective shunt capacitance across the input. However, it is clear that for many practical antenna inputs (for example some of those sketched in Fig. 1 for driving a monopole antenna above plane ground from a coaxial line), the effect of the input is too complex to be represented by a simple shunt capacitance, and a more general study of the input configuration is desirable.

It is the purpose of this paper to describe a point of view that may be applied to the input configuration of certain antennas, with examples of its use. The approach, especially that of the quasi-static method to be described, is applicable to a variety of boundary-value problems, such as those of discontinuities in wave guides.

* This work was done at the Antenna Laboratory of the University of California under Bureau of Ships Contract NObsr-39401, and constitutes a portion of a thesis submitted in partial fulfillment for the degree of Ph.D. at the University of California.

[1] G. H. Brown and O. M. Woodward, Proc. I.R.E. 33, 257–262 (April, 1945).

THE SOURCE PROBLEM IN STANDARD ANTENNA ANALYSES

In the analysis of forced oscillations on spherical[2] and spheroidal[3] antennas, Stratton and Chu summed series of the appropriate wave solutions to fit the boundary conditions of the conducting antenna with an infinitesimal gap at its center across which the driving voltage is applied. The expected infinite capacitance results, since the series for susceptance are observed to diverge. The authors, in commenting on this, have pointed out that, for a finite gap, the amplitudes of

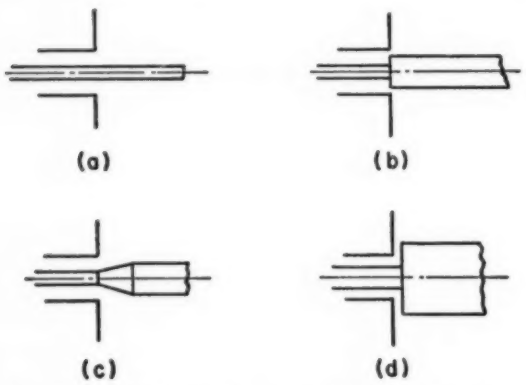


FIG. 1. Various input configurations for monopole antennas.

wave orders in the series would eventually approach zero as an order is reached great enough to give many reversals in sign in the angular function over the extent of the gap. They therefore recommend terminating the susceptance series after a finite number of terms to correspond to some actual gap. Infield[4] has made some estimates of the number of terms to retain for a given allowed error.

In Schelkunoff's analysis of the ideal biconical antenna,[5] the gap problem causes no mathematical difficulty, for, since the source applied between the tips of the two cones is a point source and not a "ring" or "slice" source as in the corresponding cylindrical dipole, there is no singularity in the admittance function. In his extension of results to thin antennas of other shapes, for example the cylindrical dipole of Fig. 2, it would appear that the original gap problem would return. Schelkunoff has a carefully outlined procedure for considering the specific gap,[6] but, even in the direct application of his perturbation solution with antenna admittance computed exactly at the origin, the infinite susceptance does not appear, and the calculated results agree well with measured values over the range of antennas for which input configuration is not important.

In order to explain the avoidance of the singularity in the Schelkunoff perturbation calculation, consider its application to the cylindrical antenna of Fig. 2. Each element of the antenna is considered first as an element of the biconical antenna passing through that radius, and the characteristic impedance computed for that point from the biconical line formula.

$$K = 120 \ln \cot(\psi/2) \cong 120 \ln[2z/\rho(z)]. \quad (1)$$

There are at least two steps in this approximation, the neglect of the finite slope of the cone passing through $\rho(z)$ in comparison with the zero slope of the cylinder, and the substitution of $2z/\rho(z)$ for $\cot(\psi/2)$. Both of these are valid for large values of $z/\rho(z)$, and these approximations enter into the limitations of the method to thin antennas. However, the approximations also fail near the origin for antennas of any size, and they fail in just such a way that the infinity of susceptance is avoided. Thus the local capacitance effects and the susceptance effects from the antenna proper are separated by this process, and the resulting admittance is satisfactory for the large number of practical cases for which the local input effects are of small importance to the antenna impedance.

For consideration of the effects of a finite gap, Schelkunoff[6] recommends that one transfer the terminating admittance obtained from his biconical line analysis only to such a radius from the origin that the approximations described above (requiring that $z/\rho(z) \gg 1$) are valid, but also small enough compared with wave-length so that quasi-static ideas may be used in this region. This admittance is then transformed to $z=0$ by uniform line equations corresponding to the biconical antenna passing through this radius. To the resulting finite admittance, one adds the difference between the capacitive susceptance of the actual cylinder, including its physically real gap, and that of the imaginary biconical antenna utilized in the above step to avoid the singularity at the origin. The method to

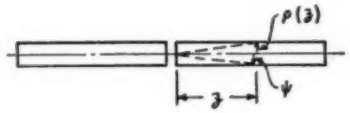


FIG. 2. Basis for Schelkunoff's formulation for other than biconical antennas.

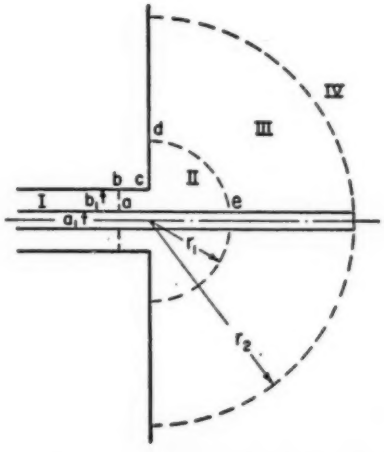


FIG. 3. Division of the antenna problem.

[2] J. A. Stratton and L. J. Chu, J. App. Phys. **12**, 236–240 (1941).
[3] L. J. Chu and J. A. Stratton, J. App. Phys. **12**, 241–248 (1941).
[4] L. Infield, Quart. App. Math. **5**, 113–132 (July, 1947).
[5] S. A. Schelkunoff, Proc. I.R.E. **29**, 493–521 (Sept. 1941).
[6] S. A. Schelkunoff, J. App. Phys. **15**, 54–60 (1944).

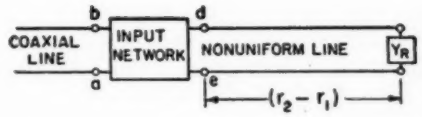


FIG. 4. Equivalent circuit for antenna of Fig. 3.

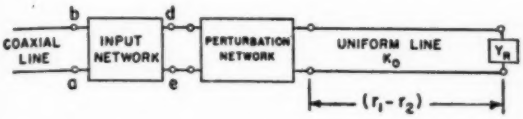


FIG. 5. Alternative equivalent circuit for the antenna of Fig. 3.

be described in this paper is in many respects an extension of this recommended procedure to other input configurations.

A third important method for the analysis of antenna impedance is that based upon approximate solutions of the integral equation developed by Hallen.[7-9] Since this analysis assumes the source as an infinitesimal gap in the cylinder across which there is a discontinuity in the scalar potential equal to the applied voltage, it would be expected to encounter the infinite susceptance difficulty described. However, it is common in these analyses to make an approximation to the angular integration in computing the vector potential, which amounts to assuming the current concentrated along the axis for purposes of computing the potential. This in effect evidently replaces the "slice" source at the gap by a point source, and so avoids the singularity as in the previous analyses.

THE POINT OF VIEW OF THIS PAPER

The point of view to be followed in this paper is: first, that the antenna system should be considered as a transducer from the feeding line to the waves in space as in Schelkunoff's analysis; secondly, that the problem should be divided into two parts, one consisting of the main body of the antenna to be analyzed substantially as in Schelkunoff's approach, and the other consisting of a region near the input, taking account of the actual input configuration. Considering specifically the monopole above ground plane, fed by a coaxial line (Fig. 3), one imaginary spherical surface is drawn through the end of the antenna at radius r_2 separating regions III and IV, and the matching of spherical wave functions across this surface is accomplished as in Schelkunoff's analysis, or by other methods. The effect of the TM modes in regions III and IV is expressed as a shunt admittance at the end of the antenna in the principal wave equivalent circuit, and this admittance is transformed to radius r_1, say by the non-uniform line formulation of Schelkunoff. All this is as in Schelkunoff's recommended procedure described above. However, at this stage, the entire region II between a reference plane $a-b$ in the coaxial line and the imaginary spherical surface $d-e$ of radius r_1 is considered as a two-terminal-pair network or transducer** whose parameters may be found either by calculation or by measurement. Once the parameters are known, and expressed in any of the possible forms (T networks, pi networks $\mathfrak{ABCD}$ constants, etc.), the admittance may be simply transformed from $d-e$ at radius r_1 to the reference $a-b$ in the coaxial line, yielding the desired information.

The reference plane $a-b$ is first selected far enough from the end of the line so that the local waves from the discontinuity have died out and the electric field is substantially radial; similarly the radius r_1 is selected far enough from the discontinuity so that the electric field lies substantially in the spherical surface $d-e$. This is the basic assumption of the approach. There could fail to be any such spherical surface if the antenna were too fat, or if it were not long enough compared to the dimensions of the input. That is, it is desired to select r_1 so that it satisfies the inequalities

$$r_1 \gg b_1 \tag{2}$$

$$r_1 \ll r_2, \tag{3}$$

where the $\gg$ is interpreted in this case only as "appreciably greater than." A reasonable selection of r_1 could thus be found easily for most antennas up to 1000 mc/sec., but not for most microwave antennas. In addition to the above inequalities, it is usually possible to choose r_1 to satisfy

$$r_1 \ll \lambda, \tag{4}$$

which is not essential, but does simplify the analysis.

The assumption that there is some region over which the electric field lines are substantially spherical, so that the analyses of the input configuration and the discontinuity at the end of the antenna may be separated, is much like that encountered in wave-guide discontinuity problems. If there are two discontinuities in a

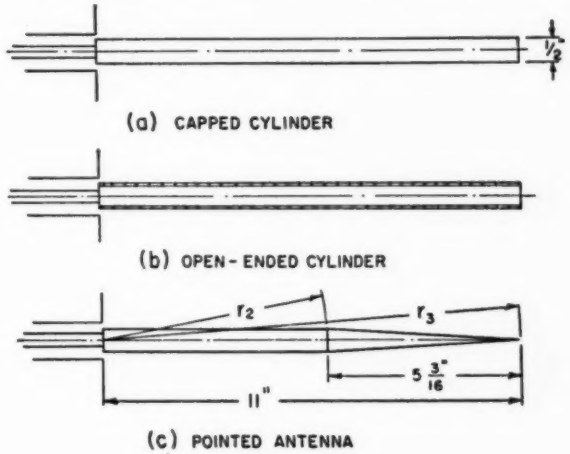


FIG. 6. The three antennas studied.

[7] E. Hallen, Nova Acta (Uppsala) **11**, 1–44 (November, 1938).
[8] C. J. Bouwkamp, Physica **9**, 609–631 (July, 1942).
[9] R. King and D. Middleton, Quart. App. Math. **3**, 302–335 (January, 1946).

** For the application of classical network theory to general electromagnetic structures, see for example, Montgomery, Dicke, and Purcell, *Principles of Microwave Circuits*, M.I.T. Rad. Lab., Vol. 8 (McGraw-Hill Book Company, Inc., New York, 1948).

wave guide far enough apart so that the local waves from each have effectively died out in some intermediate region, the dominant mode only existing in this region, the analysis of the two may proceed separately, the network for each found when alone in the guide, and the results of both found by combining by means of the transmission line equations. If they are so close that there is no such region, then they cannot be so separated.

The over-all equivalent circuit for the antenna (Fig. 4) will then consist of the input network representing region II, the non-uniform line representing the region III, and the terminating admittance Y_R representing the radiation and discontinuity effects at the end of the antenna. As will be shown later, it is possible to represent the non-uniform line by a uniform line of average characteristic impedance with a perturbation network at its input, so that the over-all equivalent circuit would then become as in Fig. 5. Of course, although the reference plane $a-b$ in the coaxial line should first be selected in the analysis far enough from the discontinuity so that there is only the principal TEM mode, and all measurements should be made at least this far from the end effect, the reference on the transmission line equivalent circuit can then be shifted at will by ordinary transmission line transformations. This will almost always be done, shifting the reference to the end of the line, since it forms the most natural reference.

THE SPECIFIC PROBLEMS STUDIED AS EXAMPLES

The point of view outlined above could be applied to a number of input configurations, such as those of Fig. 1. The particular configuration chosen for study was that of Fig. 1b, in which a $\frac{3}{4}$-in. O.D., $\frac{1}{4}$-in. I.D. coaxial line feeds a $\frac{1}{2}$-in. diameter cylindrical antenna with an abrupt step between the center conductor of the coaxial line and the antenna. This input structure was used on the three different antennas pictured in Fig. 6, each eleven inches long and $\frac{1}{2}$ inch in diameter at the base. That of Fig. 6(a) is a circular cylinder closed by a plane cap at the end, that of Fig. 6(b) is a cylinder left open at the end, and that of Fig. 6(c) is tapered in the last $5\frac{3}{16}$ inches, forming a cone with its apex at the antenna end. The input impedance for each was calculated, including the effects of the input configuration by means of the equivalent input network, and compared with measured values for a frequency range of 200 to 1000 mc/sec.

A SECOND-ORDER QUASI-STATIC CALCULATION OF THE INPUT NETWORK

In order to fulfill simultaneously the two inequalities of Eqs. (2) and (3) for the $\frac{1}{2}$-inch diameter antenna,

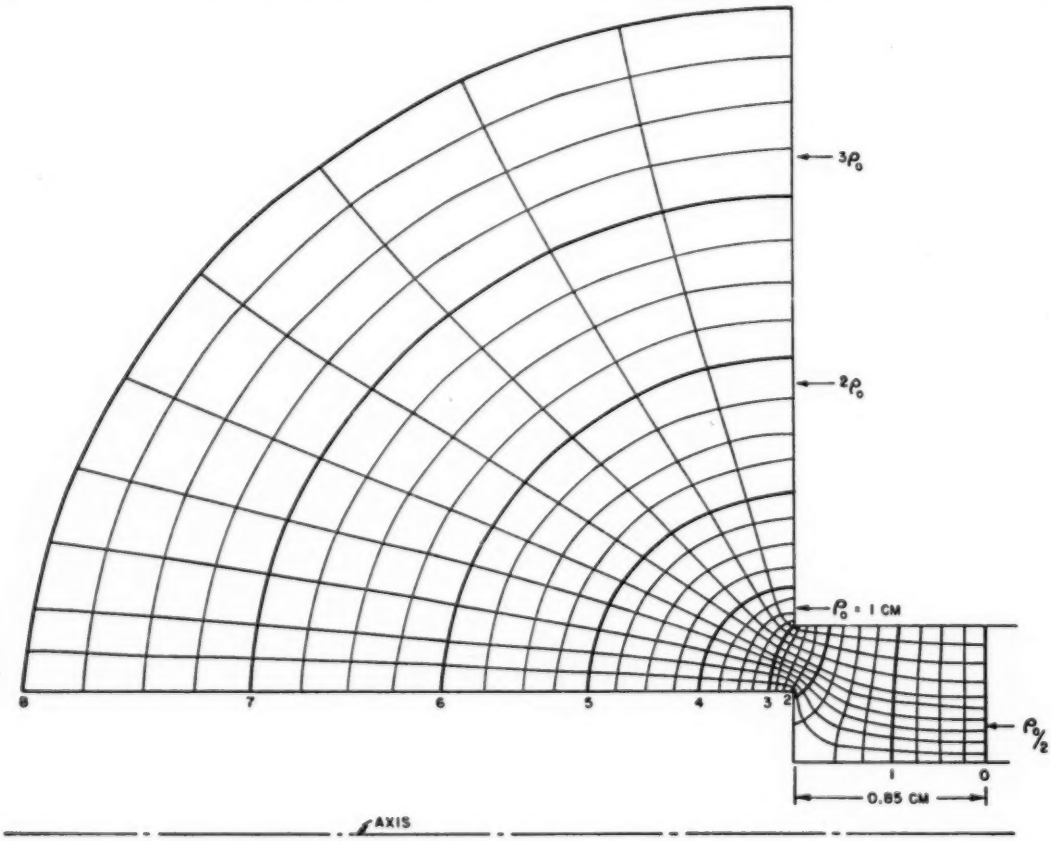


FIG. 7. Electrostatic field map of the antenna input region. Heavy flux lines indicate the divisions for the application of the second-order quasi-static analysis.

eleven inches long, fed by the $\frac{3}{4}$-inch (outer conductor inside diameter) coaxial line, the output reference sphere at radius r_1 was chosen as 3.45 cm. This yields $r_1/b=3.63$ and $r_2/r_1=8.1$, which is believed to be a satisfactory fulfillment of these inequalities. The inequality of Eq. (4) is also satisfied over the range of interest since the shortest wave-length (corresponding to 1000 mc/sec.) is 30 cm. An analysis by quasi-static methods is then appropriate for approximate determination of the input network.

A first-order equivalent circuit may be obtained from quasi-static ideas by determining for the region a series inductance L and a shunt capacitance C. The way in which the capacitance is divided between the branches of a pi (or the inductance between the arms of a T) affects the transforming properties of the network only to terms of the order ω^2LC. As is shown in Appendix A, the quasi-static ideas may be extended to give the proper division of the capacitance between the branches of the pi network for low frequencies. This is useful because the elements of the pi network actually remain substantially constant over a wide range of frequencies. For example, the experimental results show that the inductance and capacitances of the pi network representing the input configuration of the present problem change little up to 1500 mc/sec., which is considerably beyond the range of interest in this study.

The method is a quasi-static method in that it utilizes the electrostatic field map for the region, and so assumes that it is proper to approximate the wave equation by Laplace's equation. The electric field map utilized was obtained from an electrolytic tank, tilted about an axis as required for axially symmetric problems, utilizing tap water as an electrolyte, and excited by a 1000-c.p.s. signal generator. The field map obtained is shown in Fig. 7. The over-all region is subdivided by the flux tube boundaries, as indicated in Fig. 8, into regions small enough so that each of these may be represented by a symmetrical T. These are then cascaded to represent the over-all region. In the calculation of the matrix of the over-all network, terms of the order ω^2LC are retained, but higher powers of $(\omega[LC]^{\frac{1}{2}})$ are neglected. As is shown in Appendix A, results can be expressed by the pi-network of Fig. 9 with

$$L=\sum_{k=1}^{N} L_k, \tag{5}$$

$$C_a=C_n\sum_{k=1}^{N}(k-\tfrac{1}{2})\frac{L_k}{L}, \tag{6}$$

$$C_b=C_n\sum_{k=1}^{N}(N-k+\tfrac{1}{2})\frac{L_k}{L}. \tag{7}$$

The capacitance of the nth tube may be obtained from the electrostatic field map in the usual manner.

$$C=2\pi\epsilon\rho_0(N_f/N_v), \tag{8}$$

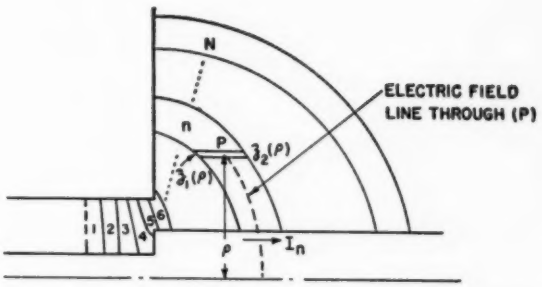
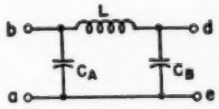


FIG. 8. Construction indicating the nth flux tube of the electrostatic field map of the region.

FIG. 9. Pi-network representing the over-all input region.



where ρ_0=radius at which side ratios of curvilinear rectangles is made unity; N_f=number of flux tubes; N_v=number of potential divisions; ϵ=dielectric constant$\cong$8.85 $\mu\mu$f/meter for air or space. The inductance of the nth tube may be obtained by graphically integrating the magnetic flux over the section of the nth tube. The magnetic field for this purpose may be related to the current in the nth tube. Thus if the line integral of magnetic field is evaluated at radius ρ passing through any point P in the nth tube (Fig. 8), it can be set equal to the current I_n, since a surface through P following an electric field line (shown dashed) would have no displacement current passing through it. So

$$H_\phi=I_n/2\pi\rho.$$

Here the current in the individual subdivision is assumed constant, and this can be satisfied by making the divisions small enough. The inductance of the nth tube is then

$$L_n=\frac{1}{I_n}\int\int \mu H_\phi d\rho dz=\frac{\mu}{2\pi}\int\int_{\text{tube section}}\frac{d\rho dz}{\rho}. \tag{9}$$

Either the ρ or z integration may be performed so that only one needs to be made graphically. Performing the z integration,

$$L_n=\frac{\mu}{2\pi}\int_{\rho\ \min}^{\rho\ \max}\frac{[z_2(\rho)-z_1(\rho)]}{\rho}d\rho, \tag{10}$$

where $z_2(\rho)$ and $z_1(\rho)$ represent the limits in z represented by the tube at a given value of radius.

For application of the method to the present problem, the region represented by the electrostatic map of Fig. 7 was divided into eight sub-regions, indicated by the heavy lines. With $N_v=10$, $N_f=4$, and $\rho_0=1$ cm, the capacitance per division from Eq. (8) was calculated as 0.222 $\mu\mu$f. The graphical integration for computation of the inductance according to Eq. (10) yielded

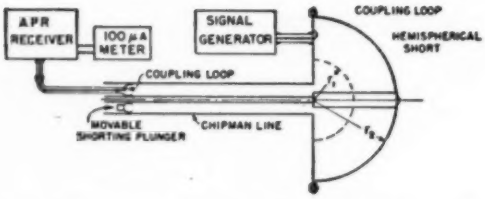


FIG. 10. Test set-up for measurement of the input region.

the following results:

$$\begin{aligned}
L_8&=4.14\times10^{-9}\text{ henrys}\\
L_7&=3.03\times10^{-9}\text{ henrys}\\
L_6&=1.88\times10^{-9}\text{ henrys}\\
L_5&=0.954\times10^{-9}\text{ henrys}\\
L_4&=0.318\times10^{-9}\text{ henrys}\\
L_3&=0.153\times10^{-9}\text{ henrys}\\
L_2&=0.858\times10^{-9}\text{ henrys}\\
L_1&=0.865\times10^{-9}\text{ henrys.}
\end{aligned}$$

Thus, using Eqs. (5)–(7)

$$L'=\sum_{k=1}^{8}L_k=12.19\times10^{-9}H, \tag{11}$$

$$C_A'=\frac{0.222}{12.19}\left[\frac{L_1}{2}+\frac{3L_2}{2}+\cdots\frac{15L_8}{2}\right]=1.250\ \mu\mu\text{f}, \tag{12}$$

$$C_B'=\frac{0.222}{12.19}\left[\frac{15L_1}{2}+\frac{13L_2}{2}+\cdots\frac{L_8}{2}\right]=0.527\ \mu\mu\text{f}. \tag{13}$$

The values of Eqs. (11)–(13) are primed because it is now desired for convenience to move the reference in the transmission line to the end of the line by subtracting out the part of the network arising from the transmission line's *principal* wave. Note that this still leaves in the network the effects of the higher order waves in the transmission line region. Writing the above network as the cascade connection of two networks, the desired one and the network representing the principal transmission line wave over the length 0.85 cm, which by usual formulas yields values of $L_0=1.87\times10^{-9}$ henry and $C_0=0.436\times10^{-12}$ Fd, the approximate elements of the desired network can be found as follows:

$$L=L'-L_0=10.32\times10^{-9}\text{ henry}, \tag{14}$$

$$C_a=\frac{L'}{L}C_A'-C_0\left(1+\frac{L_0}{2L}\right)=1.005\ \mu\mu\text{f}, \tag{15}$$

$$C_b=C_B'-\frac{L_0}{L'}\left(C_A+\frac{C_0}{2}\right)=0.340\ \mu\mu\text{f}. \tag{16}$$

EXPERIMENTAL DETERMINATION OF THE INPUT NETWORK

One of the common methods for determining the constants of a network for representation of wave-guide discontinuities utilizes the measurement of input admittances Y_{i1}, Y_{i2}, Y_{i3}, corresponding to known load admittances Y_{L1}, Y_{L2}, Y_{L3}. The admittance parameters for the network may then be calculated from the equations

$$Y_{11}=\frac{(Y_{i1}-Y_{i3})(Y_{i1}Y_{L1}-Y_{i2}Y_{L2})-(Y_{i1}-Y_{i2})(Y_{i1}Y_{L1}-Y_{i3}Y_{L3})}{(Y_{i1}-Y_{i3})(Y_{L1}-Y_{L2})-(Y_{i1}-Y_{i2})(Y_{L1}-Y_{L3})} \tag{17}$$

$$Y_{22}=\frac{(Y_{i1}Y_{L1}-Y_{i2}Y_{L2})-Y_{11}(Y_{L1}-Y_{L2})}{Y_{i2}-Y_{i1}} \tag{18}$$

$$Y_{12}^2=(Y_{11}-Y_{ik})(Y_{22}+Y_{Lk}),\quad k=1,2,3. \tag{19}$$

The known load admittances are often obtained by placing shorts on the output guide, and the input admittances measured by determining the standing wave ratio and position of a voltage node on the input guide. If losses in the system are negligible, the admittances are pure susceptances, and all Y's may be replaced by the corresponding B's (where $Y=jB$) in Eqs. (17)–(19).

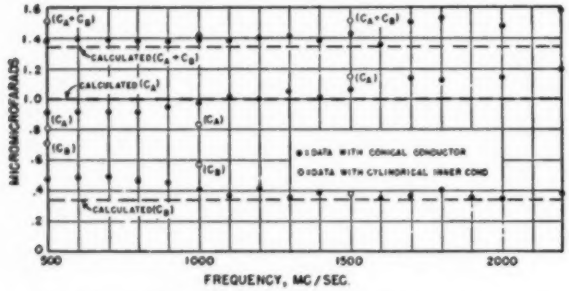


FIG. 11. Comparison of calculated and measured results for capacitances of the input network.

A similar procedure was followed in the experimental determination of the network for representation of the antenna input region in the present investigation. To produce known terminating susceptances at the reference radius r_1 chosen as the output reference of the network, hemispherical shorts to the spherical TEM mode were placed at a radius $r_2>r_1$, as indicated in Fig. 10. The positions of the current maxima on an input coaxial line were observed, giving the corresponding susceptances at the input reference. This procedure was repeated at each frequency with four different hemispherical shorts of nominal diameters 6, 9, 12, and 15 inches, the fourth pair of input-output data providing a check on the three pairs required for the calculation. Most of the measurements were made with

the inner conductor flared conically between r_1 and r_2 in order that a true spherical TEM mode might exist in this region and B_L might be computed in terms of the short position by uniform transmission line equations. It is true that this introduces a discontinuity not present on the antennas themselves, but it is believed that the effect will be small because of the smallness of the flare required. A few measurements were made for a check with the input conductor maintained cylindrical, but in these the data had to be reduced by the longer non-uniform line calculations. The test set-up for all measurements is as indicated in Fig. 10.

The admittance produced at r_1 by the length of shorted spherical TEM mode is

$$Y_L = jB_L = -j\left[\frac{1}{60 \ln \cot(\psi/2)}\right] \cot\frac{2\pi}{\lambda}(r_2 - r_1), \quad (20)$$

ψ = the cone angle, measured from the axis.

Once the admittance parameters are calculated from three pairs of input-output data by Eqs. (17)–(19), the elements of the pi-network may be found from the equations

$$\omega C_a = B_{11} - B_{12}, \quad (21)$$

$$\omega C_b = B_{22} - B_{12}, \quad (22)$$

$$\omega L = -1/B_{12}. \quad (23)$$

It has been mentioned that four hemispheres were constructed to give an additional pair of input-output data for a check, and that some check points were taken with the cylindrical inner conductors. The consistency check for the extra data was made by means of a graphical construction described by Weissfloch[10] (reactive transformation diagram). The three pairs required for the calculation from Eqs. (17)–(19) were selected as those having the greatest difference in admittance values, once unreliable points had been eliminated by the Weissfloch diagram.

The results for C_a and C_b (and for $C_a + C_b$) are shown in Fig. 11 compared with the dashed line showing the calculated result of Eqs. (15), (16). The comparison for the inductance L is shown in Fig. 12. Although agreement is not perfect, it is believed sufficient to check the principle used. Aside from the fact that the method of calculation was an approximate one, much of the difference is believed to be caused by the fact that the hemispheres could not be spun perfectly, so that they were more nearly spheroids with major and minor half-axes differing by as much as one-fourth inch. Uncertainty in reading the plunger position, although only of the order 0.1 mm, could also explain some of the deviation. A study of the equations revealed that the 0.1 mm deviation could account for as much as a three percent error in the elements of the circuit.

[10] A. Weissfloch, Hoch:tech. u. Elek:akus. **61**, 100–123 (April, 1943).

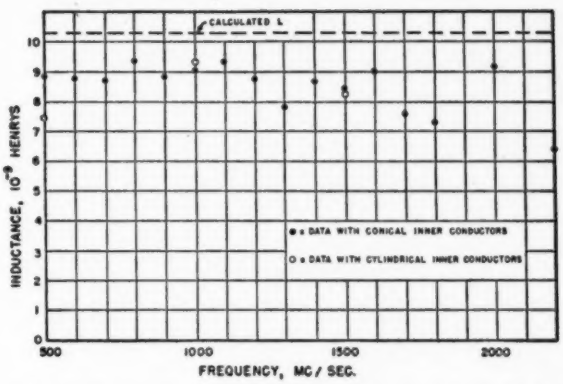


FIG. 12. Comparison of calculated and measured results for inductance of the input network.

NON-UNIFORM LINE REPRESENTATION OF THE ANTENNA

Schelkunoff's analysis of thin antennas of other than conical shape utilizes a perturbation solution which may be interpreted in terms of transformation of impedances along a non-uniform transmission system, as described earlier. As is shown in Appendix B, the Carson-Schelkunoff perturbation solution for slightly non-uniform lines may be described by stating a "perturbation" network to be connected at the input of the corresponding uniform line of average characteristic impedance and propagation constant. The $\mathfrak{A}\mathfrak{B}\mathfrak{C}\mathfrak{D}$ constants of the perturbation network are given in Eqs. (47) and (45) of Appendix B. For the antenna problem, each element of the antenna is represented by a section of the uniform spherical TEM mode, so the special forms of Eqs. (48)–(49) for a loss-free system with constant LC product at each reference on the system are applicable.

$$\mathfrak{A}_1(l) = 1 - \frac{M(l)}{K_0} \quad (24)$$

$$\mathfrak{B}_1(l) = -jN(l) \quad (25)$$

$$\mathfrak{C}_1(l) = j\frac{N(l)}{K_0^2} \quad (26)$$

$$\mathfrak{D}_1(l) = 1 + \frac{M(l)}{K_0}, \quad (27)$$

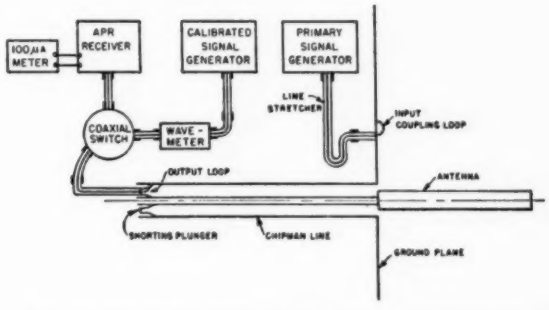


FIG. 13. Set-up for antenna impedance measurements.

where

$$M(l)=\beta\int_0^l[K_0-K(z)]\sin 2\beta z dz \tag{28}$$

$$N(l)=\beta\int_0^l[K_0-K(z)]\cos 2\beta z dz \tag{29}$$

$K(z)$ = the local characteristic impedance for a given reference z

$$K_0=\frac{1}{l}\int_0^l K(z)dz \tag{30}$$

β = phase constant, $\omega(LC)^{\frac{1}{2}}$.

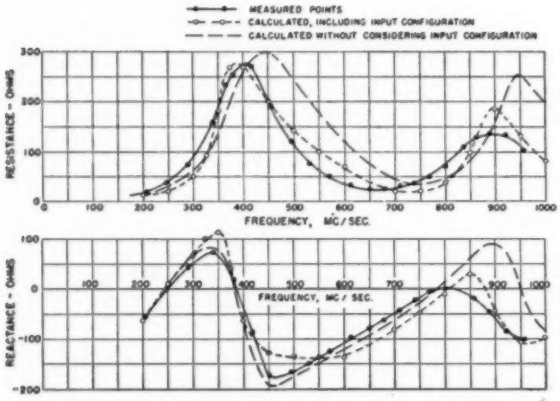


FIG. 14. Input impedance of cylindrical antenna.

For the cylindrical antenna, the equations for M, N, and K_0 given by Schelkunoff[11] must be modified slightly since it is desired to begin the nonuniform line not at the origin, but at a finite radius r_1. Thus

$$z\cong r-r_1.$$

Eq. (1) for the local characteristic impedance $K(z)$, approximated from that for the spherical TEM mode for this radius, becomes

$$K(z)\cong 60\ln\left[\frac{2r}{\rho(r)}\right]=60\ln\left[\frac{2(r_1+z)}{a}\right]. \tag{31}$$

Substitution of this value in Eqs. (28)–(30) yields

$$K_0=60\left[\frac{r_2}{l}\ln\frac{2r_2}{a}-\frac{r_1}{l}\ln\frac{2r_1}{a}-1\right] \tag{32}$$

$$M(l)=30\left\{\ln\frac{r_2}{r_1}+(1-\cos 2\beta l)\left[\frac{r_1}{l}\ln\frac{r_2}{r_1}-1\right]-\cos 2\beta r_1[Ci2\beta r_2-Ci2\beta r_1]-\sin 2\beta r_1[Si2\beta r_2-Si2\beta r_1]\right\} \tag{33}$$

$$N(l)=30\left\{\sin 2\beta l\left[\frac{r_1}{l}\ln\frac{r_2}{r_1}-1\right]+\cos 2\beta r_1[Si2\beta r_2-Si2\beta r_1]-\sin 2\beta r_1[Ci2\beta r_2-Ci2\beta r_1]\right\}. \tag{34}$$

For the pointed antenna of Fig. 6C, one transformation set may be obtained for the entire region from r_3 to r_1, or the problem may be broken into two cascaded non-uniform lines, separated at the beginning of the taper, $r=r_2$. Calculations were made both ways without much difference in the results, so the results will be given here only for the over-all antenna as a unit. The local characteristic impedance is

$$K(z)=\begin{cases}60\ln\left[\dfrac{2(r_1+z)}{a}\right] & r_1<z<r_2\\[2ex] 60\ln\left[\dfrac{2(r_1+z)l_2}{a(l-z)}\right] & r_2<z<r_3\end{cases} \tag{35}$$

Substitution of this in Eqs. (28)–(30) yields

$$K_0=60\left[\frac{r_3}{l}\ln\frac{2r_3}{a}-\frac{r_1}{l}\ln\frac{2r_1}{a}-\frac{l_1}{l}\right] \tag{36}$$

$$M(l)=[M(l)]_{cyl}+30\left[\frac{l_2}{l}(1-\cos 2\beta l)+\cos 2\beta l S_1(2\beta l_2)-\sin 2\beta l Si(2\beta l_2)\right] \tag{37}$$

$$N(l)=[N(l)]_{cyl}+30\left\{\sin 2\beta l\left[\frac{l_2}{l}-S_1(2\beta l_2)\right]-\cos 2\beta l Si(2\beta l_2)\right\}, \tag{38}$$

where $[M(l)]_{cyl}$ and $[N(l)]_{cyl}$ denote the M and N for a cylinder extending from r_1 to r_3, computed from Eqs. (33) and (34) with r_3 substituted for r_2. $S_1(x)$ is the function tabulated by Terman,[12] equal to

$$\ln x+0.5772-Ci(x).$$

THE TERMINATING ADMITTANCE

The radiation admittance at the end of the antenna, representing the effect of TM modes inside and outside the antenna region on the TEM mode equivalent circuit, may be taken from Schelkunoff's approximate formulas.[13] For a monopole, they are

$$Y_R=\frac{R_a(\beta r_2)+jX_a(\beta r_2)}{K_0^2}, \tag{39}$$

[11] S. A. Schelkunoff, *Electromagnetic Waves* (D. Van Nostrand Company, Inc., New York, 1943), p. 291.

[12] F. E. Terman, *Radio Engineers Handbook* (McGraw-Hill Book Company, Inc., New York, 1943), p. 17.

[13] See reference 11, pp. 465 and 453.

where

$$R_a(\beta r_2)=15\{2S_1(2\beta r_2)+\cos 2\beta r_2 \times[C+\ln\beta r_2-2Ci2\beta r_2+Ci4\beta r_2] +\sin 2\beta r_2[Si4\beta r_2-2Si2\beta r_2]\} \quad (40)$$

$$X_a(\beta r_2)=15\{2Si(2\beta r_2)+\sin 2\beta r_2 \times[Ci4\beta r_2-\ln\beta r_2-C]-\cos 2\beta r_2 Si4\beta r_2\}. \quad (41)$$

$$C=0.5772.$$

The approximations referred to in the development of the above limit results to relatively thin antennas. Schelkunoff has mentioned a lower limit for K_0 of around 500 ohms (for a dipole antenna) in order for the results to be applicable. The antennas studied in this report are very close to this limit.

There is also a question of a possible added admittance at the end to account for the caps or other end effect.[14] It was believed that the theory of this end effect is insufficiently understood to justify its inclusion in the calculation. The experimental results actually showed no measurable difference between the capped antenna of Fig. 6b and the open-ended antenna of Fig. 6a—a result checking similar experiments by Brown and Woodward.[1] Of course this might mean only that the two antennas had similar end effects, and not that the end effect for either is negligible. The

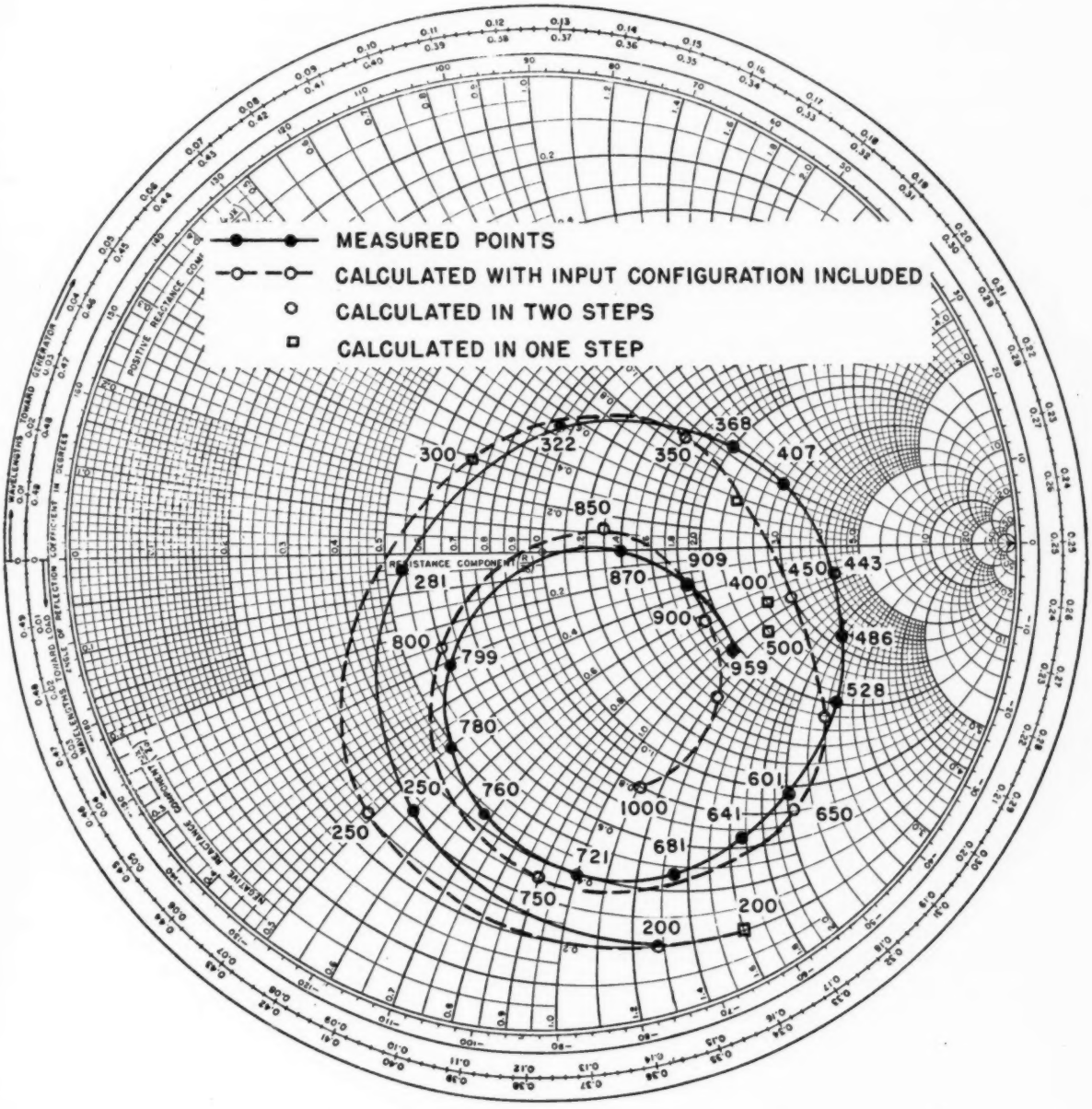


FIG. 15. Input impedance of pointed antenna referred to 66-ohm line.

[14] See reference 11, p. 465.

pointed antenna was chosen to minimize this end effect, but, since it represents a more non-uniform transmission system, the error in that stage of the analysis may be greater so that the net improvement is doubtful.

CALCULATED AND MEASURED IMPEDANCES FOR THE OVER-ALL ANTENNA

All parts of the equivalent circuit of Fig. 5 have now been described so that a calculation of the input impedance referred to the end of the antenna may be accomplished by straightforward, though lengthy, network and transmission line calculations.

The input impedance curves for the antennas sketched in Fig. 6 were measured over the range 200 to 1000 mc/sec. for comparison with the calculated values. The general set-up for the measurement is indicated in Fig. 13. The monopoles were extended horizontally from the side of a building on which there was a vertical ground plane consisting of a three-foot-square copper sheet set into a 20-foot-square section of one-fourth inch galvanized mesh. Each antenna was screwed into the one-fourth inch diameter inner conductor of the 66-ohm coaxial Chipman line.[15] Signal generator power was supplied to a semicircular loop, two inches in diameter, approximately one foot from the antenna. Received power was taken from a small loop (approximately a $\frac{1}{16}$-inch diameter semicircle) in the shorting plunger of the line by cable to an APR receiver with external meter used as a detector. The standing wave ratio (ratio of maximum current to minimum current as the shorting plunger is moved) was found by several methods, all of them utilizing comparison with an auxiliary signal generator having a calibrated attenuator. Position of the maximum current reference was also observed so that each point could be plotted directly on a Smith transmission line chart, and then translated to total impedance or admittance if desired.

Since no measurable difference was observed between the results for the capped and open-ended cylinders, only one comparison curve is shown for the cylindrical antennas. Figure 14 shows the measured curves of resistance and reactance *versus* frequency, compared with calculated values utilizing the methods of this paper, and also with curves calculated without consideration of the input configuration. These latter calculations simply extend the non-uniform line to $r=0$, as explained in the discussion of Fig. 2. The comparison of calculated and measured results for the pointed antenna is shown directly in Smith chart form in Fig. 15.

In comparing the calculated and measured curves, it is concluded that the agreement, although not perfect, is greatly improved by the consideration of the input configuration. This is true for both the cylindrical and the pointed antennas. It is believed that much of the remaining disagreement is caused by the relatively large diameter-to-length ratios of the antennas used, which make certain aspects of the thin antenna theory used subject to error. The end effect at the outer end of the antenna is also an unknown and may be important. Of course there are also possible sources of error in the measurement procedure. Among these are the finite extent and conductivity of the ground plane, the presence of reflections from nearby buildings, the finite conductivity of the measurement line, the finite size of the coupling loops, and random instabilities occasionally present in the signal generators.

[15] R. A. Chipman, J. App. Phys. **10**, 27–38 (1939).

STUDIES WITH VARIOUS INPUT CONFIGURATIONS

The study described has utilized a given input configuration and different antennas. G. W. Zeoli[16] has recently applied the approach of this paper to the study of given antennas with different input configurations. Two different input configurations were used for a $\frac{1}{2}$-inch-diameter antenna 11 inches long, and two for a $\frac{1}{4}$-inch-diameter antenna 11 inches long. There was a considerable difference between the measured results for a given antenna with the different input configurations, and the amount of this difference was predicted reasonably well by the theoretical calculations. This in some ways gives a more satisfying check of the principles introduced here than the measurements cited in the preceding section. The agreement in Zeoli's results was best for the thin antennas, checking our belief that some of the deviations in Fig. 15 arise from the application of thin antenna theory to antennas which are on the borderline of its applicability.

Work has also been reported in the literature by King and co-workers[17, 18] on antennas fed by parallel-wire lines. A terminal zone is defined there also, and, although the method of attack is somewhat different from that of this paper, equivalent networks are obtained by averaging processes for representation of the terminal zone.

APPENDIX A. A QUASI-STATIC METHOD FOR OBTAINING EQUIVALENT CIRCUITS

It is assumed that the electric field distribution is represented well enough by that obtained from a static field map, and the region to be studied is divided by the flux tube boundaries into equal flux divisions, as indicated in Fig. 8. A capacitance C_n (equal for all subdivisions) and an inductance L_n (computed from Eq. (10) for axially symmetric regions) may be found. If the subdivision is represented by a symmetrical T equivalent circuit, the $\mathfrak{A}\mathfrak{B}\mathfrak{C}\mathfrak{D}$ constants are

$$\mathfrak{A}_n=\left(1-\frac{\omega^2}{2\omega_n{}^2}\right)$$

[16] G. W. Zeoli, "Impedance of antennas with various input configurations," Report No. 157, Contract NObsr-39401, Antenna Laboratory, University of California, August 22, 1949.

[17] R. King, J. App. Phys. **20**, 832 (1949).

[18] R. King and K. Tomiyasu, Proc. I.R.E. **37**, 1134 (1949).

$$\mathfrak{B}_n = j\omega L_n\left(1-\frac{\omega^2}{4\omega_n{}^2}\right)$$

$$\mathfrak{C}_n = j\omega C_n$$

$$\mathfrak{D}_n = 1-\frac{\omega^2}{2\omega_n{}^2},$$

where

$$\omega_n{}^2 = \omega^2 L_n C_n.$$

The N subdivisions are now cascaded and the overall $\mathfrak{ABCD}$ constants obtained by multiplying the individual matrices together. Multiplying from the right, the product of the last two matrices is taken, neglecting powers of (ω/ω_n) higher than the second.

$$\mathfrak{A}_{N-1,N} = 1-\frac{\omega^2 C_n}{2}(L_N+3L_{N-1})$$

$$\mathfrak{B}_{N-1,N} = j\omega(L_N+L_{N-1})$$

$$\mathfrak{C}_{N-1,N} = 2j\omega C_n$$

$$\mathfrak{D}_{N-1,N} = 1-\frac{\omega^2 C_n}{2}(3L_N+L_{N-1}).$$

Repetition of this process with the succeeding adjacent matrices reveals the law of formation, as follows:

$$\left.\begin{aligned}
\mathfrak{A} &= 1-\frac{\omega^2 C_n}{2}\sum_{k=1}^{N} L_k[2(N-k)+1] \\
\mathfrak{B} &= j\omega\sum_{k=1}^{N} L_k \\
\mathfrak{C} &= NC_n \\
\mathfrak{D} &= 1-\frac{\omega^2 C_n}{2}\sum_{k=1}^{N}(2k-1)L_k.
\end{aligned}\right\} \tag{42}$$

The above may be shown to be the constants for a pi-network as shown in Fig. 9 with L, C_a, and C_b given by Eqs. (5) and (7). The same formulas are applicable to a two-dimensional region, except that the formulas for inductance of the subdivisions (10) must be modified.

APPENDIX B. NETWORK FORMULATION OF NON-UNIFORM LINE THEORY

For a slightly non-uniform transmission system with local impedance per unit length given as $Z(z)=Z_0+\hat{Z}(z)$, and the local admittance per unit length given as $Y(z)=Y_0+\hat{Y}(z)$, Schelkunoff[19] has given the zero-order solution in his Eqs. (11-7), and the first-order solution, obtained by a perturbation calculation, in Eqs. (12-1). The zero- and first-order solutions may be added and put in matrix form as follows:

$$\begin{bmatrix} V(0) \\ I(0) \end{bmatrix} = \begin{bmatrix} \mathfrak{A}(l) & \mathfrak{B}(l) \\ \mathfrak{C}(l) & \mathfrak{D}(l) \end{bmatrix}\begin{bmatrix} V(l) \\ I(l) \end{bmatrix}, \tag{43}$$

where

$$\left.\begin{aligned}
\mathfrak{A}(l) &= [1-F_2(l)]\cosh\gamma_0 l \\
&\quad +[F_1(l)+F_3(l)]\sinh\gamma_0 l \\
\mathfrak{B}(l) &= K_0\{[1-F_2(l)]\sinh\gamma_0 l \\
&\quad +[F_1(l)+F_3(l)]\cosh\gamma_0 l\} \\
\mathfrak{C}(l) &= \frac{1}{K_0}\{[F_3(l)-F_1(l)]\cosh\gamma_0 l \\
&\quad +[1+F_2(l)]\sinh\gamma_0 l\} \\
\mathfrak{D}(l) &= [1+F_2(l)]\cosh\gamma_0 l \\
&\quad +[F_3(l)-F_1(l)]\sinh\gamma_0 l
\end{aligned}\right\} \tag{44}$$

$$\left.\begin{aligned}
F_1(l) &= \frac{1}{2}\int_0^l\left[\frac{\hat{Z}(z)}{K_0}-K_0\hat{Y}(z)\right]\cosh 2\gamma_0 z\,dz \\
F_2(l) &= \frac{1}{2}\int_0^l\left[\frac{\hat{Z}(z)}{K_0}-K_0\hat{Y}(z)\right]\sinh 2\gamma_0 z\,dz \\
F_3(l) &= \frac{1}{2}\int_0^l\left[\frac{\hat{Z}(z)}{K_0}+K_0\hat{Y}(z)\right]dz.
\end{aligned}\right\} \tag{45}$$

K_0 and γ_0 are the average characteristic impedance and propagation constant, respectively.

The matrix of Eq. (43) defined by (44), may be recognized as the product of two matrices.

$$\begin{bmatrix} \mathfrak{A}(l) & \mathfrak{B}(l) \\ \mathfrak{C}(l) & \mathfrak{D}(l) \end{bmatrix} = \begin{bmatrix} \mathfrak{A}_1(l) & \mathfrak{B}_1(l) \\ \mathfrak{C}_1(l) & \mathfrak{D}_1(l) \end{bmatrix} \times \begin{bmatrix} \cosh\gamma_0 l & K_0\sinh\gamma_0 l \\ \dfrac{1}{K_0}\sinh\gamma_0 l & \cosh\gamma_0 l \end{bmatrix}, \tag{46}$$

where

$$\left.\begin{aligned}
\mathfrak{A}_1(l) &= 1-F_2(l) \\
\mathfrak{B}_1(l) &= K_0[F_1(l)+F_3(l)] \\
\mathfrak{C}_1(l) &= \frac{1}{K_0}[F_3(l)-F_1(l)] \\
\mathfrak{D}_1(l) &= 1+F_2(l).
\end{aligned}\right\} \tag{47}$$

[19] See reference 11, Sections 7.11 and 7.12.

Thus the over-all transformation network for the non-uniform line may be broken into a section of uniform line of characteristic impedance K_0 and propagation constant γ_0, with a transforming network with constants given by Eqs. (47) in cascade with its input.

The non-uniform systems met with in this study are specialized in that they are loss free, and also have a constant LC product for each point along the line. The following may then be shown:

$$\frac{\hat{Z}}{K_0}+K_0\hat{Y}=0$$

$$\frac{\hat{Z}}{K_0}-K_0\hat{Y}=2j\omega(L_0C_0)^{\frac{1}{2}}\left[\frac{K(z)}{K_0}-1\right].$$

Substitution of these relations in Eqs. (61) gives

$$F_1(l)=\frac{N(l)}{jK_0}$$
$$F_2(l)=\frac{M(l)}{K_0} \qquad (48)$$
$$F_3(l)=0,$$

where

$$M(l)=\beta\int_0^l[K_0-K(z)]\sin 2\beta z dz$$
$$N(l)=\beta\int_0^l[K_0-K(z)]\cos 2\beta z dz. \qquad (49)$$

The $\mathfrak{A}_1$, $\mathfrak{B}_1$, $\mathfrak{C}_1$, $\mathfrak{D}_1$, constants of Eqs. (47) then become as given in the main body of the paper, Eqs. (24)–(27).

Microwave Techniques for the Measurement of the Dielectric Constant of Fibers and Films of High Polymers

T. M. Shaw and J. J. Windle
Western Regional Research Laboratory, Albany, California*
(Received March 27, 1950)

A resonant cavity method is described for the determination of the dielectric constant of polymeric solids in the form of fibers, films, and thin cylinders. Methods for investigating electrical anisotropy are outlined. In the case of fibers, for the electric field parallel to the fiber axis, the dielectric constant is determined directly. For the electric field perpendicular to the fiber axis, the problem arises of calculating the dielectric constant of the fiber from that of an air-fiber mixture. For the latter purpose the Lichtenecker equation was found to be applicable. Illustrative measurements of the dielectric constant of wool, Nylon, and Cellophane are presented and discussed briefly.

INTRODUCTION

THE development of sources of microwave energy has made possible the extension of dielectric-constant studies on solid polymers to frequencies in the region 3000 to 30,000 Mc (1 Mc = 10^6 cycles per second) or more where new facts concerning the structure of polymers are to be expected. The keratins represent a group of polymers for which dielectric studies in the microwave range should be especially informative. Dielectric constants for wool keratin[1] and for horn keratin[2] have been reported for frequencies between about 500 c.p.s. and 13 Mc. At 13 Mc the dielectric constant was found to be 4.2. This value is considerably greater than the square of the index of refraction for visible light ($\sim$2.25) and has been interpreted as an indication of atomic or ionic polarization.[1] This should lead to a large variation in the dielectric constant with frequency in the microwave region. In order to investigate this possibility, suitable methods were sought for making dielectric constant measurements on the keratins in the microwave region.

Since the keratins of greatest interest are available only in limited forms, for example wool fibers, the generally used cavity resonator[3] and wave-guide procedures,[4,5] which employ relatively large accurately machined specimens, are not directly applicable. For this reason a modified cavity resonator technique has been developed for use with dielectric specimens in the form of fibers, films, and thin cylinders. The procedures and apparatus are described in the present paper. They have been used exclusively in the 3000-Mc frequency region but appear to be suitable for use at higher frequencies with only minor modifications.

* Bureau of Agricultural and Industrial Chemistry, Agricultural Research Administration, United States Department of Agriculture. Report of a study made under the Research and Marketing Act of 1946.

[1] J. Errera and H. S. Sack, Ind. Eng. Chem. **35**, 712 (1943).
[2] G. King, Trans. Faraday Soc. **43**, 601 (1947).
[3] Horner, Taylor, Dunsmuir, Lamb, and Jackson, J. Inst. Elec. Eng. **93**, 53 (1946).
[4] S. Roberts and A. von Hippel, J. App. Phys. **17**, 610 (1946).
[5] For a general discussion of methods for making dielectric constant measurements at microwave frequencies see R. M. Redheffer, *Technique of Microwave Measurements* (McGraw-Hill Book Company, Inc., New York, 1947), p. 561.

THEORY

Figure 1 represents a cavity resonator. The cavity is circular in cross section and is closed with plane parallel ends perpendicular to the axis of the cylinder. As is well known, when a cavity is excited to resonance at the frequency f_0, in the fundamental or TM_{010} mode, the electric field in the cavity is directed parallel to the cylinder axis. The electric field possesses a single maximum at the axis and falls to essentially zero at the cylinder wall.

The usefulness of such a resonator for dielectric constant measurements derives from the fact that when a dielectric specimen is placed in the cavity the resonance frequency is changed; such change in frequency can be related quantitatively to the dielectric properties of the specimen.

The theory of the use of the resonator for the case of a smooth circular cylinder low loss[6] dielectric specimen placed along the cavity axis is well known.[3] For a specimen so small in diameter that insertion of the specimen in the cavity results in a negligible change in the electric and magnetic field distribution, the resonant frequency is given by the equation,

$$f=f_0-(\epsilon-1)f_0b^2/0.539a^2. \qquad (1)$$

In this equation f is the frequency of the resonator containing the dielectric specimen, f_0 is the frequency of the empty resonator, ϵ is the dielectric constant of the specimen, b is the radius of the dielectric cylinder, and a is the radius of the cavity resonator.

The frequency change (f_0-f) to be expected when a single wool fiber of a length equal to that of the cavity is used as a dielectric specimen is easily obtained by means of Eq. (1). Assuming as typical values, $\epsilon\sim4$, $b\sim10^{-3}$ cm, $a=4.5$ cm, $f_0=3000$ Mc, it is found that (f_0-f) ~0.001 Mc. Although this small frequency change can be detected readily with standard microwave equipment, it is too small for accurate measurement purposes. Since changes in the temperature of the cavity and in the humidity of the air in the cavity can produce relatively large frequency changes it is necessary to work with specimens consisting of several thousand fibers. In this way only moderately constant temperature and humidity are required.[7]

A cylindrical bundle containing several thousand wool fibers is about 0.15 cm in radius. The theoretical calculations of Horner and co-workers[3] show that if Eq. (1) is applied to a cylinder of dielectric of about this size and with a dielectric constant of six or less, the frequency change will be accurate to within one percent. This is true because for a cylinder of dielectric of this diameter the variation of the intensity of the electric field across the cylinder cross section is negligible.

For convenience in calculations Eq. (1) may be expressed in terms of the volume of the dielectric specimen and the volume of the cavity resonator by replacing b^2 by its equivalent $v_s/\pi l$ and a^2 by its equivalent $v_c/\pi l$. The resultant equation is

$$f=f_0-(\epsilon-1)f_0v_s/0.539v_c. \qquad (1a)$$

In this equation it is understood of course that the volume v_s is that of a specimen extending the entire length of the cavity.

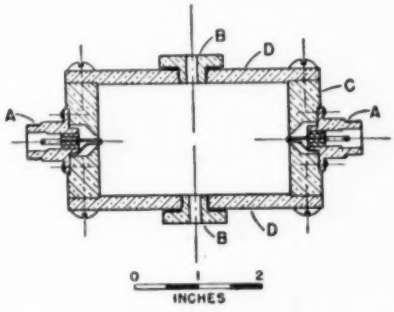


FIG. 1. Cross-sectional view of cylindrical cavity resonator. *A*, Type *N* cable connectors; *B*, plugs to permit specimen to be inserted along axis of resonator; *C*, cylinder wall, *D*, end plates perpendicular to cylinder axis.

Equation (1) and thus also (1a) was derived for a dielectric specimen consisting of a single circular cylinder. Since a bundle of fibers only approximates a circular cylinder in shape, tests were made to determine whether errors were introduced by applying Eq. (1a) to a specimen composed of a number of circular cylinders. Glass was chosen as a convenient test material rather than a polymer since it was desired to avoid possible effects due to anisotropy and to water absorption. Single glass cylindrical rods were prepared with radii ranging from about 0.02 to 0.05 cm. These rods were inserted in the cavity individually and in groups and the frequency change (f_0-f) determined. The results are shown in Table I. Constancy of the ratio $(f_0-f)/b^2$ shows that the frequency change (f_0-f) is a linear function of the volume of glass contained in a specimen irrespective of whether the glass is in the form of a single rod or a group of rods. Calculations showed that the external radius of the largest bundle of rods considered was within the limit for the size of a single cylinder set by Horner[3] if an accuracy of one percent or better is to be expected. Table I also shows that dielectric constants calculated from the various measurements are in good agreement. The statistical error (standard deviation) is slightly less than one percent which, considering the approximations made in the derivation, is all that can be expected from Eq. (1a) for specimens of this size and dielectric constant.

These results demonstrate that the exact form of the

[6] The equations given below apply strictly only to a hypothetical dielectric with a zero loss tangent, i.e., $\tan\delta=0$. Their application to the materials considered in the present paper is justified because $\tan\delta$ does not exceed 0.05 and only a negligible error is introduced. For a complete discussion see reference 5, p. 568.

[7] For the particular 3000-Mc cavity resonator used a one-degree rise in temperature produces a decrease in the resonant frequency of 0.058 Mc. A change in relative humidity from 25 percent to 50 percent will cause a frequency decrease of about 0.03 Mc.

boundary of the dielectric specimen is not important provided the specimen is confined to a region where the electric field is essentially constant as discussed above. Since the shape of the specimen is not important, it was assumed and subsequently verified with Teflon films and rods that the dielectric specimen may have any of the forms A, B, C, E, or F of Fig. 2 provided only that the specimen is continuous over the entire length of the cavity resonator and be composed of elements of solid the boundaries of which are tangential to the electric field. For all such specimens Eq. (1a) can be used to calculate the dielectric constant. The volume of the specimen is conveniently and accurately determined from its mass and density. This procedure is especially convenient for fiber specimens where the small size and large number of the fibers in a specimen makes accurate measurements of the fiber volume very difficult.

TABLE I. Summary of measurements of dielectric constant of Pyrex glass (No. 7740) rods.

Specimen numbers	Square of radius of rod b^2	Change in frequency (f_0-f)	$(f_0-f)/b^2$	Dielectric constant
	$(cm)^2$	Mc		
1	27.10×10^{-4}	3.82	1.410×10^3	4.66
2	26.80	3.83	1.430	4.70
3	22.40	3.18	1.420	4.68
4	21.80	3.06	1.405	4.64
5	19.05	2.66	1.395	4.62
6	15.95	2.24	1.405	4.64
7	12.52	1.79	1.430	4.70
8	9.36	1.32	1.410	4.65
9	9.12	1.32	1.448	4.73
10	6.76	0.972	1.439	4.73
11	4.68	0.686	1.464	4.79
1, 3, 4	71.30	10.15	1.425	4.68
6, 7, 8	37.83	5.41	1.430	4.70
9, 10, 11	20.56	2.96	1.440	4.73
2, 7, 8, 11	53.36	7.61	1.428	4.69
		Average	1.425 ± 0.018	4.69 ± 0.04

Specimens of type E are of interest when it is desired to investigate the electrical anisotropy in polymer films.[8] The general treatment of E has been discussed. It remains only to remark that by rolling the film into a thin cylinder with the selected direction in the film parallel to the axis of the cylinder it is possible to determine the dielectric constant of the film with the electric field parallel to the direction chosen. By rolling the film into cylinders with the axis in other directions anisotropy can be investigated.

The type of specimen represented by D of Fig. 2 is perhaps the most important that must be considered. A solenoidal specimen of fibers of this type is used to determine the dielectric constant of the fibers with the electric field perpendicular to the fiber axis. Such measurements when combined with those on parallel bundles of fibers may be used to investigate the electrical anisotropy of fibers.[8]

A solenoidal specimen is obtained by winding fibers in the form of a helix with very small pitch about a central core. The core is made from a material of low dielectric constant and dielectric loss in order to minimize the correction to the frequency change to be made for the core. Polystyrene is satisfactory from this viewpoint but it is not satisfactory when it is necessary to heat the fibers such as is required in some instances to remove absorbed water. It was because of the latter limitation that quartz was usually used in the present work.

The dielectric constant which results when Eq. (1a) is applied to a solenoidal specimen is that of an *air-solid* mixture. The volume of the specimen v_s used in Eq. (1a) is calculated from the external diameters of the solenoid and its core and the length of the solenoid. The length is, of course, the same as that of the resonator. From these measurements the dielectric constant of the air-solid mixture (fibers plus air) is calculated. The justification for this procedure lies in the fact that the mixture of fibers and air occupies an essentially uniform electric field and the frequency changes due to the core and to the fiber-air mixture are additive.

To deduce the dielectric constant of the solid material (fibers) from that of the air-fiber mixture recourse must be had to a mixture law. For this purpose consideration has been given to only the equations of Lorentz[9] and of Lichtenecker,[10] for dielectric mixtures. For a two-component system the Lorentz equation is

$$(\epsilon_m-1)/(\epsilon_m+2)=v_f(\epsilon_\perp-1)/(\epsilon_\perp+2)+v_a(\epsilon_a-1)/(\epsilon_a+2). \quad (2)$$

In this equation ϵ_m is the dielectric constant of the air-fiber mixture, $\epsilon_\perp$ is the dielectric constant of the fiber substance measured with the electric field directed perpendicular to the fiber axis, ϵ_a is the dielectric constant of air and v_f and v_a are, respectively, the volume of the fiber substance and of air in 1 cc of the fiber-air mixture.

Similarly, for a two-component mixture the Lichtenecker equation may be written,

$$\log\epsilon_m=v_f\log\epsilon_\perp+v_a\log\epsilon_a, \quad (3)$$

and the symbols have the same meanings stated for Eq. (2).

Owing to the well-recognized theoretical limitations of mixture equations for dielectrics it was desired to ascertain only on a purely empirical basis whether one of these equations would give satisfactory values for $\epsilon_\perp$. The validity of the equations was tested by comparing the value of $\epsilon_{||}$ determined for glass fibers with $\epsilon_\perp$ for the same fibers as calculated by means of the mixture equa-

[8] A preliminary discussion of methods for measurement of electrical anisotropy in polymers at microwave frequencies was presented by the authors at the Seattle meeting of the American Physical Society, June 27–29, 1949; T. M. Shaw and J. J. Windle, Phys. Rev. 76, 586 (1949).

[9] H. A. Lorentz, *Theory of Electrons* (Teubner, Berlin, 1916), second edition, p. 147.

[10] K. Lichtenecker, Physik. Zeits. 27, 115 (1926).

tions from measurements performed on solenoidal specimens. For glass fibers $\epsilon_{\parallel}$ should equal $\epsilon_{\perp}$ since well-annealed glass is isotropic. Table II summarizes the experimental data and calculations for a number of specimens of glass fibers. The table shows that the mean value of $\epsilon_{\perp}$ calculated with the aid of the Lichtenecker equation agrees with $\epsilon_{\parallel}$ within about two percent. This is considered to be satisfactory because of the likelihood of errors which can arise in the measurement of the external diameter of the solenoid. An error of one percent in diameter of the solenoid can cause an error of about two percent in $\epsilon_{\perp}$. It is significant in this regard that a low value of $\epsilon_{\perp}$ is to be associated with a low value in the diameter of the solenoid. This is the most likely direction for error since it is not possible to define exactly the outer boundary of the air-fiber mixture.

The values of $\epsilon_{\perp}$ obtained by means of Eq. (3) are higher than $\epsilon_{\parallel}$ by about 7 percent. Because the disagreement between $\epsilon_{\parallel}$ and $\epsilon_{\perp}$ obtained with the Lichtenecker equation is no greater than the experimental error it was adopted as satisfactory for calculating the dielectric constant of the fibers from measurements on an air-fiber mixture in the form of a solenoid. The table also shows that ϵ_m for the air-fiber mixture as well as $\epsilon_{\perp}$ is independent of the size of the solenoidal specimen. This constitutes further experimental proof that the electric field is essentially uniform over the cross section occupied by the solenoidal dielectric specimens.

EXPERIMENTAL

A cross-sectional view of the cavity resonator is shown in Fig. 1. The cavity dimensions are almost identical with those for a similar cavity described by Horner.[3] The present cavity was equipped with standard type *N* coaxial cable connectors which terminate in small coupling loops. These loops are inserted into the cavity through holes in the cylinder wall. The plane ends of the cavity are equipped with removable plugs in order to insert dielectric specimens. For non-hygroscopic specimens the removable plugs are simply drilled with holes large enough to accommodate the specimen. The effect of the holes and the specimen contained within them on the resonant frequency of the cavity is very small.[11] For hygroscopic specimens the removable plugs are designed to hold the specimen in place in the cavity and to allow the cavity to be evacuated.

When the cavity was used at temperatures other than room temperature a correction was made for the change in resonant frequency caused by expansion or contraction of the cavity. The correction required was determined experimentally to be 0.058 Mc per degree C by measuring the resonant frequency of the cavity over the temperature range 0 to 100°C.

A block diagram of the microwave equipment used to determine the frequency of the resonant cavity is shown in Fig. 3. The klystron oscillator is frequency modulated

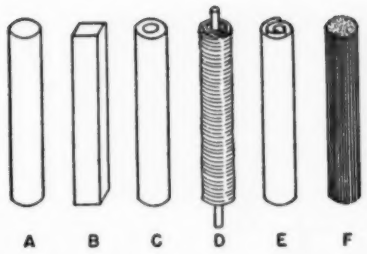


FIG. 2. Dielectric specimens. *A*, solid circular cylinder; *B*, solid rectangular bar; *C*, hollow circular cylinder; *D*, fiber solenoid; *E*, thin film wound into circular cylinder; *F*, bundle of fibers.

by means of a saw-tooth wave derived from the oscilloscope sweep. The signal from the klystron is attenuated and fed simultaneously to the cavity resonator and to a wave meter. The rectified output of the cavity is combined with the rectified output from the wave meter in a balanced mixer transformer. The resultant signal is amplified and displayed on the oscilloscope. The frequency of the oscillator is adjusted until it is being swept over the frequency of the cavity resonator. When this is done, a typical resonance curve is obtained. The wave meter is adjusted until it is tuned to the same frequency as the cavity as indicated by the appearance of a double-humped pattern on the oscilloscope. By proper adjustment of the wave-meter output the pattern obtained on the oscilloscope is a very sensitive indication of the equality of frequency of the cavity and wave meter. The general usefulness of such patterns in the tuning of resonators has been discussed.[12] The same procedure is followed to determine the frequency f_0 of the empty cavity resonator and the frequency, f, of the cavity resonator when it contains a dielectric specimen.

The wave meter used to determine the frequency difference (f_0-f) is especially designed to facilitate the determination of small frequency differences. The wave meter consists of a section of rectangular wave guide tuned by means of movable plungers in the ends. Operation of the wave meter in the TE_{107} mode results in a relatively large displacement in plunger position for

TABLE II. Application of Lichtenecker and Lorentz equations to a mixture of glass fibers and air.

	$\epsilon_{\parallel}$ (glass fibers) $=5.99\pm0.05$				
Specimen number	Volume of fiber-air mixture	Volume of fibers	Dielectric constant of mixture	$\epsilon_{\perp}$ calculated Lichtenecker equation	$\epsilon_{\perp}$ calculated Lorentz equation
	(cc)	(cc)	ϵ_m		
1	14.67×10^{-2}	7.10×10^{-2}	2.35	5.86	6.42
2	20.20	9.15	2.24	5.87	6.52
3	35.15	16.10	2.25	5.88	6.39
4	14.92	6.60	2.20	5.98	6.43
5	10.20	4.83	2.31	5.86	6.44
6	17.68	8.51	2.34	5.85	6.42
7	23.88	11.75	2.40	5.93	6.49
			Average	5.89 ± 0.05	6.44 ± 0.05

[11] J. C. Slater, Rev. Mod. Phys. **18**, 441 (1946).

[12] R. L. Sproull and E. G. Linder, Proc. I.R.E. **34**, 305 (1946).

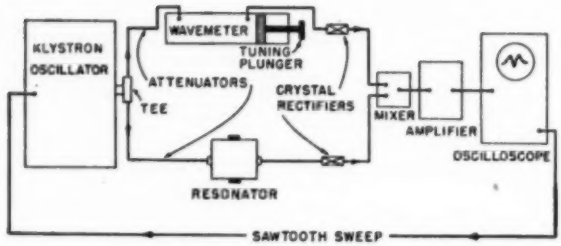


FIG. 3. Block diagram of microwave apparatus.

a small frequency difference (f_0-f). The calibration for the wave meter was calculated from the dimensions of the wave guide. Rotation of the screw used to drive the tuning plunger by one turn changes the frequency by 2.756 Mc. The sensitivity of the measuring system is such that a change in frequency of 0.003 Mc can be detected easily.

Verification of the calculated calibration of the wave meter was accomplished indirectly by measuring the dielectric constants of a number of materials for which the dielectric constants at 3000 Mc are given in the literature.[13] Figure 4 is a graph comparing the values obtained by the present method with the values taken from the literature.[13] With the exception of the data for glasses, the values given in the literature and those determined by the methods described in this paper agree to within about one percent. For the glasses deviations up to about five percent were obtained. In view of the discrepancy in the low frequency dielectric constants for glasses reported by different sources, it was assumed that the deviations greater than one percent observed in the microwave dielectric constants are due to differences in composition of the specimens rather than to errors in measurement. This assumption is supported by the measurements summarized in Table I which show that variations in the dielectric constant for various sizes of samples of a given glass are much smaller than the discrepancy between the literature values and the present determinations. For the data in Table I the standard deviation is less than one percent.

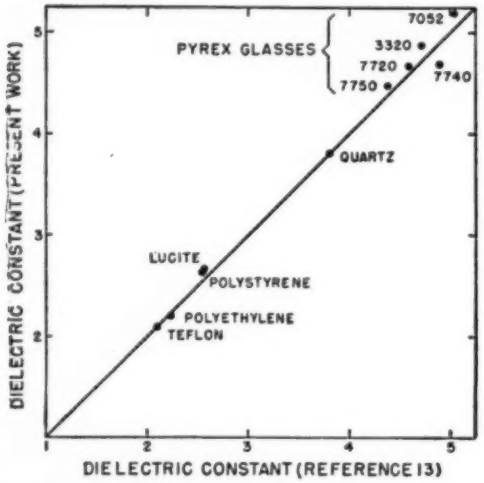


FIG. 4. Comparison between dielectric constants measured by methods described in present paper and data of A. von Hippel (reference 13).

[13] A. Von Hippel, Tables of Dielectric Materials, Vols. I and II, Laboratory for Insulation Research, M.I.T. Cambridge, Massachusetts, NDRC Division 14, Report No. 266 and 282.

TYPICAL RESULTS

The dielectric constants of a number of natural and synthetic polymers have been measured by the methods outlined in this paper. The measurements presented below have been selected to illustrate a number of the types of specimen which have been studied.

The dielectric constant of dry wool fibers has been determined for the electric field parallel to the fiber axis ($\epsilon_{\parallel}$) and for the electric field perpendicular to the fiber axis ($\epsilon_{\perp}$). The fibers were arranged into bundles and solenoids as illustrated in Fig. 2 and dried at 100°C in a vacuum oven. The results are shown in Table III. The values of the dielectric constant for the individual specimens agree to within about plus or minus 3 percent. This variation is believed to be due to errors in the dimensions of the specimens. The wool fibers are not uniform and it is difficult to prepare specimens that are well defined geometrically. If methods can be perfected to prepare more uniform specimens, the variation in the dielectric constant can probably be reduced substantially.

The differences between the values of $\epsilon_{\parallel}$ and $\epsilon_{\perp}$ for wool shown in Table III are sufficiently great compared to the statistical errors that it appears likely wool is electrically anisotropic at 3000 Mc. This is interesting in view of the anisotropy reported for another keratin[2] (horn) at much lower frequencies and in view of the apparent lack of anisotropy in horn at 3000 Mc.[8] The validity of the conclusion that wool is electrically anisotropic is of course based on the applicability of the

TABLE III. Dielectric constant of dry wool at 3000 Mc.

Specimen number	Dielectric constant $\epsilon_{\parallel}$*
1	4.01
2	3.66
3	3.86
4	3.75
5	4.03
6	3.97
	Average 3.88±0.15
Specimen number	Dielectric constant $\epsilon_{\perp}$*
7	4.40
8	4.31
9	4.38
10	4.56
	Average 4.41±0.11

* $\epsilon_{\parallel}$ = dielectric constant for electric field parallel to fiber axis. $\epsilon_{\perp}$ = dielectric constant for electric field perpendicular to fiber axis.

TABLE IV. Dielectric constant of Nylon fibers at 3000 Mc (fibers stretched about 300 percent before measurements, moisture content 2.0±0.5 percent).

Specimen number	Dielectric constant ϵ_{11}*
1	3.18
2	3.14
3	3.09
4	3.11
	Average 3.13±0.04

* ϵ_{11} = dielectric constant for electric field parallel to fiber axis.

TABLE V. Dielectric constant of dry Cellophane film (electric field parallel to plane of film).

Specimen	Dielectric constant
1	4.05
2	4.06
3	4.04
4	4.03
5	4.00
6	4.06
7	4.01
8	4.06
9	4.10
10	4.02
	Average 4.04±0.03

Lichtenecker equations to mixtures of fibers and air as discussed earlier. If these equations are not valid for wool, then the conclusion of an anisotropy for wool would have to be modified. Further discussion of the anisotropy in wool and other related polymers will be reserved for discussion at a later time.

Table IV contains values of the dielectric constant (ϵ_{11}) of Nylon fibers. The agreement between the individual measurements of (ϵ_{11}) is better than was obtained for wool. This is probably a result of the fact that the Nylon fibers measured were in the form of a uniform continuous filament. This made it possible to prepare more uniform specimens than was possible with wool.

In Table V results are presented which illustrate an application to measurement of the dielectric constant of a polymer film. The material in this case is Cellophane. The ten specimens for which data are given were cut from a single sheet and rolled to form cylinders in the manner illustrated by E of Fig. 2. The individual values of the dielectric constant agree to within about plus or minus one percent, which indicates a high degree of uniformity in a single sheet of film. Similar measurements on specimens from a second sheet of film gave an average value for the dielectric constant of 3.89 compared to the value of 4.04 obtained with the first sheet. This difference in the dielectric constant of the two sheets probably represents a real difference in the Cellophane.

From the results presented it may be concluded that the resonant cavity method may be used to advantage for the determination of the dielectric constants of polymers either in the form of thin cylinders, fibers, or films. It thus appears to be especially useful in the study of the dielectric properties of natural polymers most of which are available only in highly specialized or limited forms such as, for example, fine fibers. The method appears to be capable of yielding accurate values of the dielectric constant even when used with rather small amounts of material.

A more complete discussion of the dielectric constants of wool and related polymers obtained by the methods described above will be given in a subsequent paper to be submitted to this journal.

Otto A. Beeck

It is with sincere regret that we announce the death on July 5, 1950 of Dr. Otto A. Beeck, Associate Editor of the *Journal of Applied Physics*.

Determination of Residual Stresses of Quenching Origin in Solid and Concentric Hollow Cylinders from Interferometric Observations

A. W. SÁENZ
Applied Mathematics Branch, Mechanics Division, Naval Research Laboratory, Washington, D. C.
(Received August 10, 1949)

A theoretical procedure for the determination of the stresses of quenching origin in solid and concentric hollow cylinders has been derived. The method relies on interferometric data and the assumptions discussed in Section I, whose validity for the technically important case of glass is supported by a number of experimental investigations.

The results should be of interest for the analysis of stresses in quenched glass cylinders, so great in magnitude that the usual cutting techniques are precluded, because of the explosive characteristics of highly stressed objects of this material.

I. INTRODUCTORY REMARKS

AN optical method for the determination of the pointwise distribution of stresses of thermoelastic or residual type due to quenching in transparent initially isotropic spheres and infinite cylinders was developed recently at the University of Michigan, and presented in an earlier paper.[1] In the former case, it was proved that the residual stresses could be measured by means of a Babinet compensator, provided the photoelastic law was valid for these stresses, while in the latter it was necessary to introduce special assumptions concerning the mechanism of quenching. To avoid the use of such hypotheses, a theoretical method for the determination of residual stresses of quenching origin in solid and concentric hollow cylinders from interferometric observations alone will be developed in this article. This extension of the use of the usual interferometric procedures[2] to three-dimensional stress fields will complete the work on spheres and cylinders mentioned previously. The method should be of particular interest for the case of glass.

The considerations in the following pages rest on three assumptions: (a) The directions of the principal stresses are assumed to be parallel to the cylindrical coordinates ρ, φ, and z. (b) The stress system resulting from quenching is assumed to be radially symmetrical, i.e., it is independent of φ and z. (c) The Maxwell-Neumann stress-optic law holds, i.e.,

$$\begin{aligned} n_P - n_0 &= C_1 P + C_2(Q+R), \\ n_Q - n_0 &= C_1 Q + C_2(P+R), \end{aligned} \tag{1}$$

where P and Q are secondary principal stresses in the plane of the wave front, R is the stress normal to the wave front, n_P and n_Q are refractive indices for plane polarized waves with directions of vibration parallel to P and Q, respectively. As to the meaning of n_0 in Eq. (1), a few words are necessary at this point. In the case of stresses below the elastic limit induced by mechanical, thermal, or other means, the initial index of refraction of the material involved is equal to the final index, when the external forces are removed. This is the ordinary case treated in photoelasticity, and experiment shows that (1) is valid if n_0 is identified with the initial refractive index of the substance under examination. However, when residual stresses are being studied photoelastically, the index of refraction of the material is frequently history-dependent, the case of glass furnishing a good illustration of this statement.[3] To avoid these difficulties, the stress-optic law (1) will be assumed in this paper with the proviso that n_0 is now a constant on the same footing as C_1 and C_2. An experimental method to find this constant will be given in Section III.

How valid are the assumptions (a)–(c) for infinite cylinders which have been symmetrically quenched? It is clear that (a) and (b) hold for objects of this shape under the quenching conditions mentioned previously, from symmetry arguments. As to the third assumption, it is realized that this is a highly controversial subject at the present time. Notwithstanding, a number of experimental results strongly support the validity of the photoelastic law for residual stresses of quenching origin in glass, for the case when the initial temperature is low enough so that plastic deformations can be neglected. For example, the conditions of static equilibrium are obeyed by stresses in glass, calculated by photoelastic means,[4] and it is also known that the birefringence of quenched objects of this material is affected by cutting and goes to zero at the surface of the cut, which is a good indication that the optical effect can be expressed as a function of the stress alone. Of course, it will be necessary to carry out a more exhaustive experimental investigation before the above assertions can be generally accepted.

[1] R. C. O'Rourke and A. W. Sáenz, Quart. App. Math. (to be published). The distinction between thermoelastic and residual stresses due to quenching, which is made in that article and the present one, is that the former are transient, because of the fact that the initial temperature is below the softening temperature, while the latter are permanent in character and are produced by rapid cooling from temperatures above the softening point.

[2] See, for example, H. Favre, Rev. d'optique 8, 193–213, 241–261, 289–307 (1929).

[3] Tool, Tilton, and Saunders, J. Research Nat. Bur. Stand. 38, 519 (1947).

[4] F. L. Everett and K. A. Parsons, "Progress reports on stresses in insulators," Eng. Research Inst., University of Michigan (January to June, 1948).

II. DETERMINATION OF σ_ρ, σ_φ, AND σ_z IN INFINITE SOLID AND CONCENTRIC HOLLOW CYLINDERS FROM INTERFEROMETRIC DATA

If a symmetrically quenched infinite cylinder is immersed in a liquid whose index of refraction matches its average index and a plane polarized beam of light is

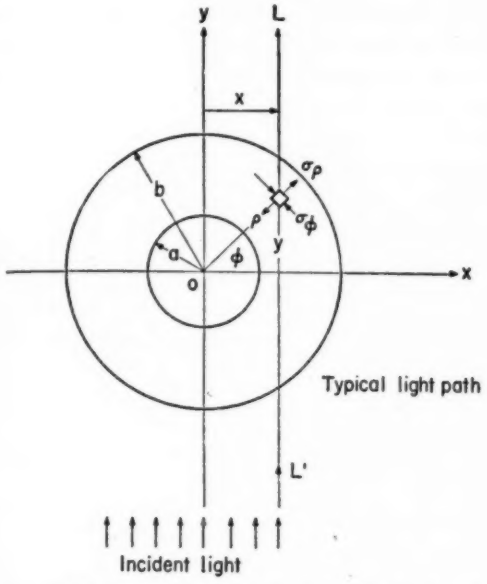


FIG. 1(a) Light normally incident on cylinder (z axis perpendicular to plane of paper).

made to fall normally on it, say along the y axis, as shown in Fig. 1(a), assumption (a) implies:

$$\left.\begin{aligned}\sigma_x&=\sigma_\rho\cos^2\varphi+\sigma_\varphi\sin^2\varphi\\ \sigma_y&=\sigma_\rho\sin^2\varphi+\sigma_\varphi\cos^2\varphi\\ \tau_{xz}&=0\\ \sigma_z&=\sigma_z\end{aligned}\right\}, \tag{2}$$

which means that the following identification can be made:

$$P=\sigma_z,\quad Q=\sigma_x,\quad R=\sigma_y \tag{3}$$

(nothing new would result by putting, instead of (3), $P=\sigma_x$, $Q=\sigma_z$) and, therefore, the directions of the principal stresses do not rotate in the plane of the wave front, so that the usual methods of calculating the phase retardations are correct in this case.

The absolute phase retardations R_P and R_Q along P and Q are, from (1), (2), and (3),

$$\left.\begin{aligned}R_P(x)&=\int_{-[b^2-x^2]^{\frac{1}{2}}}^{[b^2-x^2]^{\frac{1}{2}}}n_p dy\\ &=2\int_x^b\{C_1\sigma_z+C_2(\sigma_\rho+\sigma_\varphi)\}\frac{\rho d\rho}{[\rho^2-x^2]^{\frac{1}{2}}}\\ &\quad+2n_0[b^2-x^2]^{\frac{1}{2}}\end{aligned}\right\}, \tag{4a}$$

$$\left.\begin{aligned}R_Q(x)&=\int_{-[b^2-x^2]^{\frac{1}{2}}}^{[b^2-x^2]^{\frac{1}{2}}}n_Q dy\\ &=2\int_x^b\{C_1(\sigma_\rho\cos^2\varphi+\sigma_\varphi\sin^2\varphi)\\ &\quad+C_2(\sigma_z+\sigma_\rho\sin^2\varphi+\sigma_\varphi\cos^2\varphi)\}\frac{\rho d\rho}{[\rho^2-x^2]^{\frac{1}{2}}}\\ &\quad+2n_0[b^2-x^2]^{\frac{1}{2}}\end{aligned}\right\}, \tag{4b}$$

where x is the distance of a given ray LL' [Fig. 1(a)] from the origin 0, measured along the x axis. Equations (4) hold in the interval

$$a\leqq x\leqq b,\quad\text{with}\quad 0\leqq a<b, \tag{4'}$$

where a is the inner radius of the cylinder, taken equal to zero in the case of solid cylinders. Thus, there is no need to distinguish formally between solid and concentric hollow cylinders.

For the sake of completeness, a description of the measurement of $R_P(x)$ and $R_Q(x)$ will be sketched at this point, although it has no essential experimental novelty. A Mach-Zehnder interferometer, as modified by Favre[5] for photoelastic purposes, may be employed

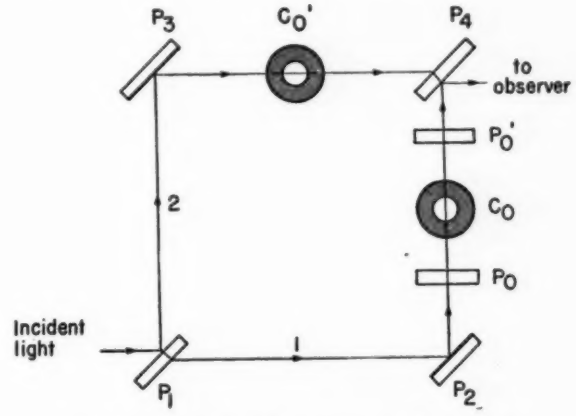


FIG. 1(b) Illustration of measurement of $R_P(x)$ and $R_Q(x)$ by a Mach-Zehnder interferometer.

[Fig. 1(b)]. The plates P_2 and P_3 are full-silvered mirrors, while P_1 and P_2 are half-silvered. The cylinder whose residual stresses are to be determined is placed at C_0, between the half-wave plates P_0 and P_0', in the path of ray 1, and a compensating cylinder of the same size and material is put at C_0', in the path of ray 2. The incident light is plane polarized and has its direction of vibration perpendicular to the plane of the paper [Fig. 1(b)], and the plates P_0 and P_0' permit one to rotate the plane of vibration so that it is parallel to either P or Q, i.e., to the z and x axes, respectively, at C_0

[5] See reference 2, pp. 245–253, where a careful description of the measurement of the absolute retardations along the P and Q axes is given.

and emerges vertical to the plane of the paper after having traversed P_0'. Thus the absolute retardations $R_P(x)$ and $R_Q(x)$ may be measured by Favre's method regardless of the fact that one is dealing with a three-dimensional stress system, because the secondary principal axes do not rotate along the light path for normal incidence.

Regarding $R_P(x)$ and $R_Q(x)$ as known for the interval (4'), it will be demonstrated that σ_ρ, σ_φ, and σ_z can be uniquely determined for all points of the cylinder examined, by inversion of the integral equations (4).

Subtracting (4b) from (4a) and putting $\cos\varphi = x/\rho$, one obtains:

$$R_P(x)-R_Q(x)\equiv r(x)$$

$$=2C\int_x^b\left\{(\sigma_z-\sigma_\varphi)+(\sigma_\varphi-\sigma_\rho)\frac{x^2}{\rho^2}\right\}\frac{\rho d\rho}{[\rho^2-x^2]^{\frac{1}{2}}} \quad (5)$$

where $C=C_1-C_2$ is the relative photoelastic constant and $r(x)$ is the relative retardation at the point x [Fig. 1(a)].

One can eliminate σ_φ from (5) by employing the equation of equilibrium

$$(d\sigma_\rho/d\rho)+[(\sigma_\rho-\sigma_\varphi)/\rho]=0. \quad (6)$$

If one carries out this substitution, integrates (5) partially to get rid of the term in $d\sigma_\rho/d\rho$, and demands that σ_ρ vanish at $\rho=b$, on account of the free boundary, one obtains:

$$r(x)=2C\int_x^b\frac{\sigma_z(\rho)\rho d\rho}{[\rho^2-x^2]^{\frac{1}{2}}}. \quad (7)$$

This is an integral equation of Abel type, index $\frac{1}{2}$,[6] whose only continuous solution is given by

$$\rho\sigma_z(\rho)=-\frac{1}{\pi C}\frac{d}{d\rho}\int_\rho^b\frac{r(x)xdx}{[x^2-\rho^2]^{\frac{1}{2}}},$$

which presents difficulties at $\rho=0$. To arrive at a more convenient expression for σ_z one integrates the last equation by parts[7] and uses the fact that $r(b)=0$, as can be seen from (5), so that the solution of the integral equation (7) becomes

$$\sigma_z(\rho)=-\frac{1}{\pi C}\int_\rho^b\frac{[dr(x)/dx]}{[x^2-\rho^2]^{\frac{1}{2}}}dx. \quad (8)$$

Digressing for a moment at this point, it can be seen that (7) expresses the remarkable fact that, as far as measurements of relative retardation for normal incidence of plane polarized light are concerned, a symmetrically quenched infinite cylinder acts as if only the axial stress σ_z were present. An interesting by-product of (7) and (8) is that σ_z can be determined by the Babinet compensator method, without recourse to more elaborate interferometric procedures.

Returning to the main discussion, one can also find similar expressions for σ_ρ and σ_φ in terms of optical data. With this purpose in mind, (4a) can be written as follows, using (7):

$$\left.\begin{aligned}\int_x^b\{\sigma_\rho(e)+\sigma_\varphi(e)\}\frac{\rho d\rho}{[\rho^2-x^2]^{\frac{1}{2}}}\\ =\frac{1}{2C_2}\{R_p(x)-2n_0[b^2-x^2]^{\frac{1}{2}}\\ -(C_1/C_2)r(x)\}\equiv F(x)\end{aligned}\right\}, \quad (9)$$

where $F(x)$ is known from interferometric observations.

Since (9) is again an Abel integral equation of index $\frac{1}{2}$, one can invert it uniquely in the same way as was done for (7), with the result

$$\sigma_\rho(\rho)+\sigma_\varphi(\rho)=-\frac{2}{\pi}\int_\rho^b\frac{dF(x)/dx}{[x^2-\rho^2]^{\frac{1}{2}}}. \quad (10)$$

From (6) and (10) one has, by repeated application of Dirichlet's formula for the inversion of order in repeated integrals,

$$\left.\begin{aligned}\sigma_\rho(\rho)=\frac{2}{\pi}\int_\rho^b\frac{S(x)dx}{x^2[x^2-\rho^2]^{\frac{1}{2}}}\\ \text{where}\\ S(x)\equiv\int_x^b dF(x')/dx'\end{aligned}\right\}. \quad (11)$$

Finally, from (6) and (11) it follows that

$$\sigma_\varphi(\rho)=\frac{2}{\pi}\frac{d}{d\rho}\left\{\rho\int_\rho^b\frac{S(x)dx}{x^2[x^2-\rho^2]^{\frac{1}{2}}}\right\}. \quad (12)$$

Equations (8), (11), and (12) furnish the principal stresses for solid and concentric hollow cylinders once n_0 has been measured. The integrations can be carried out by numerical or analytical methods, but it is felt that the latter are preferable for the most accurate results.[8] Since current interferometric methods in photoelasticity are capable of measuring absolute retardations down to a few percent of a wave-length, the present method seems quite promising with respect to accuracy. It is

[6] E. T. Whittaker and G. N. Watson, *Modern Analysis* (The Cambridge University Press, New York, 1945), p. 229.

[7] This and the succeeding partial integrations and inversions of the order of integration of repeated integrals are justified in virtue of the fact that the absolute retardations, $R_P(x)$ and $R_Q(x)$, as shown by physical considerations, are continuous in the closed interval (4') and possess derivatives which are also continuous in this interval, except at $x=b$, where they have singularities $0\{(b^2-x^2)^{-\frac{1}{2}}\}$.

[8] Analytical procedures were employed in conjunction with the fitting of the retardation curves by polynomials, using least squares, in a similar connection, with very satisfactory results. See Everett, Parsons, O'Rourke, and Sáenz, "Progress reports on stresses in insulators," Eng. Research Inst. Publ. (December, 1947), and also reference 4.

planned to carry out systematic experimental investigations with symmetrically quenched, unannealed glass cylinders to confirm this point.

III. MEASUREMENT OF n_0

Since it is necessary to know n_0 to calculate the stress distribution σ_ρ, σ_φ, and σ_z from the above integral formulas, a method for determining it will be presented at this point.

The following procedure can be used to measure n_0. One rotates the cylinder about the x axis by an angle α, so that the light strikes it obliquely, and determines the absolute retardations of the rays located in the central section S, as shown in Fig. 2. One also finds the value of the absolute retardations for waves whose planes of vibration coincide with the yz plane and are incident normally on the cylinder, as in Fig. 1(b). This information determines n_0 completely.

To prove this statement, one proceeds in the following way. In section S

$$\left.\begin{aligned} P' &= \sigma_z \cos^2\alpha + \sigma_\rho \sin^2\alpha \\ Q' &= \sigma_\varphi \\ R' &= \sigma_z \sin^2\alpha + \sigma_\rho \cos^2\alpha \end{aligned}\right\}, \tag{13}$$

where P' and Q' are the secondary principal stresses in the plane of the wave front in section S (Fig. 2), and R' is the stress normal to the wave front. Since it can be shown that no rotation of the secondary principal axes occurs along the light path LL' in S, the absolute retardations $R_P^{(\alpha)}$ and $R_Q^{(\alpha)}$ (the superscripts refer to the angle α of oblique incidence) corresponding to plane polarized waves whose planes of vibration are parallel to P' and Q', respectively, are given by

$$\left.\begin{aligned} R_{P'}^{(\alpha)} &= 2\sec\alpha\left\{\int_a^b [C_1(\sigma_z\cos^2\alpha+\sigma_\rho\sin^2\alpha) \right. \\ &\quad \left. +C_2(\sigma_\varphi+\sigma_z\sin^2\alpha+\sigma_\rho\cos^2\alpha)]d\rho+n_0(b-a)\right\} \\ R_{Q'}^{(\alpha)} &= 2\sec\alpha\left\{\int_a^b [C_1\sigma_\varphi+C_2(\sigma_\rho+\sigma_z)]d\rho \right. \\ &\quad \left. +n_0(b-a)\right\} \end{aligned}\right\}, \tag{14}$$

where the factor $\sec\alpha$ arises from the fact that the light paths are longer for oblique than for normal incidence.

The terms in (14) containing σ_φ vanish because of the relation

$$\int_a^b \sigma_\varphi(\rho)d\rho = b\sigma_\rho(b) - a\sigma_\rho(a), \tag{15}$$

which follows from (6). The right-hand side is zero for both solid and hollow cylinders since the boundaries are free from external stresses. From (14) and (15),

$$\left.\begin{aligned} n_0 = \frac{1}{2C(b-a)}\{&C_2R_{P'}^{(0)}(\cot^2\alpha-1)+[-C_2R_{P'}^{(\alpha)} \\ &+(C_1+C_2)R_{Q'}^{(\alpha)}\sin^2\alpha]\cot\alpha\csc\alpha\} \end{aligned}\right\}, \tag{16}$$

where $R_{P'}^{(0)}$ is the absolute retardation through the section S (Fig. 2) along the P' direction for normal incidence ($\alpha=0$).

Equations (8), (11), (12), and (16) permit one to determine σ_ρ, σ_φ, and σ_z at all interior and boundary points of symmetrically quenched solid cylinders and concentric hollow cylinders from purely optical data.

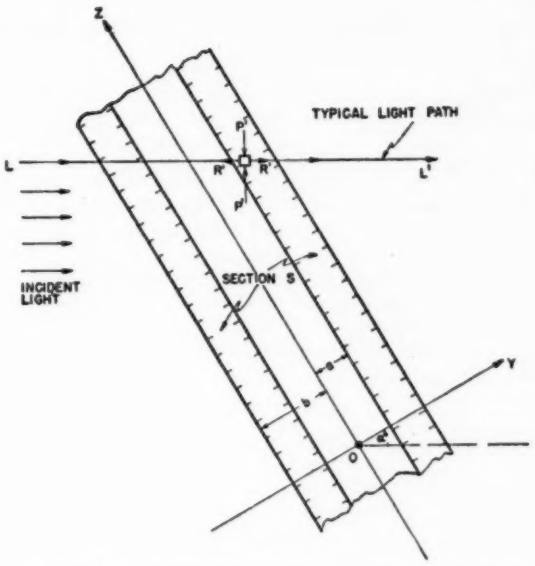


FIG. 2. Oblique incidence on central section S of cylinder (x axis and Q' perpendicular to plane of paper).

IV. CONCLUDING REMARKS

The method presented in the previous sections should be of particular interest for the analysis of residual stresses in glass objects which are so highly quenched that their explosive characteristics would prevent the use of cutting techniques for the measurement of these stresses. It also appears promising in the study of transient thermal stresses in cylinders, particularly for those cases when the temperature differences are so large that thermal and elastic constants of the material vary considerably, and the radiation at the surface does not follow the simple Newton law of cooling, so that the Stefan-Boltzmann law must be employed. Such thermal stresses cannot be determined readily by present photoelastic techniques.

Some Crossover Properties in the Electron Immersion Objective

L. JACOB*
The University, Manchester, England
(Received January 31, 1950)

It was earlier established that the lens field between the cathode and crossover can be represented by a simple function $V=A\sinh kz$, so that the main properties of the paraxial "crossover" (or minimum section of the beam) can be determined as a first approximation in terms of "A" and "k." The distance of the crossover works out at $8\pi/3k\sqrt{3}$; its potential is $A\exp(8\pi/3\sqrt{3})$; and the angle of the beam is simply proportional to "k." The radius of the crossover is then $2/k\exp[-(2\pi/3\sqrt{3})]\cdot\tan\theta$ for electrons of initial velocity "A," where θ is the angle of emission within the paraxial limit. The differential analyzer was also used to obtain the correct solution for the trajectory equation from the field plots for different systems, and shows that the simple theory gives an adequate approximation for design purposes.

I. INTRODUCTION

IT was only with the advent of the study of electron motions in E. S. Fields that their property of crossing over to form a minimum beam section was recognized. The optical counterpart has no practical significance, since focusing of the object is the chief aim, and in a sense it was hardly ever considered. In their observations on the focused spots of television C.R. tubes, Maloff and Epstein[1] concluded that the "object" which was imaged on the screen was a "crossover" produced by the first lens (immersion objective).

This is brought about in the following way. The electrons, which leave the cathode at all angles and with a range of initial velocities, are focused by the immersion lens and produce an image of the cathode. This is usually enlarged and inverted. The "principal rays" however, cross the axis at a point nearer the cathode; the plane perpendicular to the axis at this point contains the "crossover." The envelope of the beam at the crossover gives the minimum cross section of the beam, which is ultimately focused into the small spot on the screen.

The immersion objective consisted of an oxide emitting cathode C, a modulating electrode M, and an accelerator A (Fig. 1). The distance between C and M was varied keeping the position of A fixed, and the field distribution determined in the electrolytic tank. It has been shown[2] that a simple empirical formula $V=Ae^{kz}$ approximately describes the relation between the potential V at any point on the axis and the distance "z" from the cathode, for the region up to and somewhat beyond the modulating electrode, which was explored in the tank. Using this, an approximate theory of crossover properties is presented in what follows, which can be helpful in design. This indicates the importance of the geometrical arrangement of the electrodes.

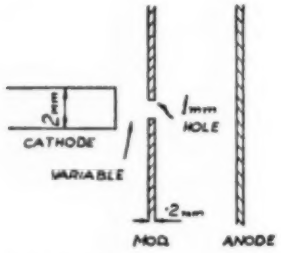


FIG. 1. Immersion objective.

II. THE CATHODE/CROSSOVER DISTANCE

For the purposes of the subsequent analysis, the simple relation $V=Ae^{kz}$ has been used. Here "k" is a geometrical scale factor which varies from geometry to geometry, while "A" is a voltage scale factor, which, for any one geometry and fixed field form, varies correspondingly with the over-all voltage applied to the system. In order to satisfy the boundary conditions at the cathode ($V=0$), the distribution in the region between the cathode and the nearest point of the experimental plot should be represented by $V=A(e^{kz}-e^{-kz})$. The extent of the linear part of the field given by the last expression is then assumed small in comparison with the exponential portion Ae^{kz} which includes the crossover region.

Electrons which are emitted normal to the cathode in the linear field remain normal till they enter the exponential region. Here, they experience the focusing effect. By assuming that the function $V=Ae^{kz}$ applies to this region, the electrons which leave the cathode at $z=0$ must enter the exponential field with a velocity of "A" volts. For paraxial systems this has a value of about 0.5 for a 30-v blackout, and is thus of significance in practice.

We begin with the paraxial form of the trajectory equation:

$$d^2r/dz^2+\tfrac{1}{2}(V'/V)dr/dz+\tfrac{1}{4}(V''/V)r=0 \qquad (1)$$

and, substituting $V=Ae^{kz}$, this yields

$$r=C\exp\{[-k/4+(k\sqrt{3}/4)\cdot i]z\}+D\exp\{[-k/4-(k\sqrt{3}/4)i]z\} \qquad (2)$$

as the complete solution. This gives the height "r" of the ray at any point

$$r=e^{-(k/4)z}\{M\cos(k\sqrt{3}/4)z+N\sin(k\sqrt{3}/4)z\}. \qquad (3)$$

To evaluate the constants M and N, we must put in the

* Now at University of Liverpool, England.
[1] I. G. Maloff and D. W. Epstein, *Electron-Optics in Television* (McGraw-Hill Book Company, Inc., New York, 1938).
[2] P. A. Einstein and L. Jacob, Phil. Mag. 39, 20 (1948).

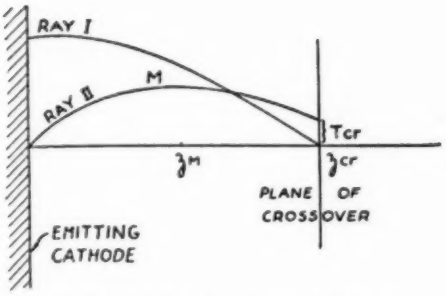


FIG. 2. Rays defining crossover and its plane.

boundary conditions. An electron starting at the top of the emitting area and moving parallel to the axis is deviated by the field so as to cut the axis at the crossover (Fig. 2); this gives one focal plane of the immersion objective. Now put $r=a$, the radius of the emitting area when $z=0$; hence $M=a$. Also $(dr/dz)_{z=0}=0$; hence

$$N=a/\sqrt{3}; \tag{4}$$

$$\therefore\quad r=ae^{-(k/4)z}\{\cos(k\sqrt{3}/4)z+(1/\sqrt{3})\sin(k\sqrt{3}/4)\cdot z\}. \tag{5}$$

Now, let this trajectory meet the axis in the crossover plane, at a distance z_{cr} from the cathode. Here, $r=0$.

$$\therefore\quad \cos(k\sqrt{3}/4)\cdot z_{cr}+(1/\sqrt{3})\sin(k\sqrt{3}/4)\cdot z_{cr}=0; \tag{6}$$

$$\therefore\quad z_{cr}=8\pi/3k\sqrt{3}. \tag{7}$$

Thus, the crossover distance from the cathode varies inversely as "k," the geometrical scale factor.

Actually, the value of the crossover distance determined from (7) will be a limiting value, since it corresponds to electrons with initial speeds of about 0.5 v. The most probable crossover corresponding with the most probable velocity of about 0.1 volt will lie somewhat closer to the cathode, but the difference is only some five percent and may be overlooked for this type of system. This "chromatic aberration" effect may however be important in systems where there is a wide spread in z_{cr} or of initial velocity; its effect will be most marked when crossover diameters are to be derived.

III. THE BEAM ANGLE AND POTENTIAL AT THE CROSSOVER

The immersion objective has the form of a "thick lens," i.e., consists of a series of contiguous refracting layers, so that an electron ray is continuously refracted all through the field. The direction in which it emerges and which defines the angle of the beam, as measured in a field free space outside the lens, is usually very different from the angle φ at which it crosses the axis at the crossover. The angle φ is derived by differentiating (5) and putting

$$z=z_{cr}=8\pi/3k\sqrt{3};$$

$$\therefore\quad \tan\varphi=-(ak/2)\exp(-2\pi/3\sqrt{3}). \tag{8}$$

This gives the angle at the crossover, in terms of the height at which the ray left the cathode normally, and the "k" of the field. For a constant emitting area of radius "a," the angle at the crossover is directly proportional to the factor "k" which defines the geometry of the system. It is, of course, also proportional to the diameter of the coated area.

For a given geometry, the beam angle is thus constant, independent of the total voltage on the system, because the factor "A" does not determine its value. This is a well known experimental law which is used in the Electron-optical voltmeter.[3]

The potential on the axis at the crossover is simply derived by substituting the value of z_{cr} from (7) in the equation $V=Ae^{kz}$ for the axial distribution. This yields a value $A\exp(8\pi/3\sqrt{3})$, which is thus proportional to A; hence, A is some constant fraction of the crossover voltage. Under the approximation used, it is also the initial velocity with which the electrons enter the exponential part of the field, as stated previously. If the total voltage on the system is increased "n" times, that at the crossover also increases "n" times, and the ratio of crossover potential to final potential is not altered. Hence, no change would be expected in the magnification of an image of the crossover, when the total voltage on the tube was varied in any way, apart from the effect on the space charge. Use can be made of this fact by finding

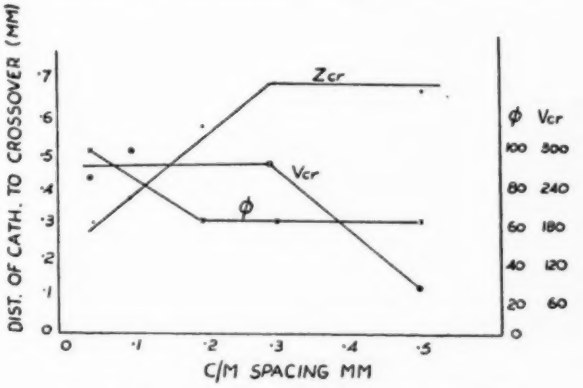


FIG. 3. Effect of spacing on crossover parameters.

the point at which the relationship no longer holds; the control of crossover size by space charge then steps in, and the magnitude involved can be assessed.

The variation of these crossover properties with C/M spacing is shown in Fig. 3. These were derived from the field plots for the different systems. They include crossover distance z_{cr}, angle at the crossover φ and its potential V_{cr} calculated from the above formulae using the constants determined from each geometry. The data has been converted to correspond with the case of a vacuum tube system run under the condition of constant anode voltage for each system. It will be seen that the potential at the crossover remains nearly constant except for the largest spacing. Also, the beam angle decreases and appears to reach a limiting value of 60° at

[3] L. Jacob, J. Inst. Elec. Eng. **91**, 512 (1944).

this spacing. In an actual sealed off tube, the largest angle at the crossover is about 100°. In practice, using a fluorescent disk to measure the angle, the measured angles are only about 20°, i.e., one-fifth the calculated values. This arises from the fact that the angle is measured in a field-free space at the final anode voltage, while the angle at the crossover is situated in a space at a very much lower potential. Care should be taken to allow for this difference, when using measured values in any calculations which have to be made.

IV. THE EFFECT OF MODULATION ON CROSSOVER PROPERTIES

The field plotting technique can be readily applied to follow through the effect of modulating the beam, by applying negative potentials, V_m, to the modulator,

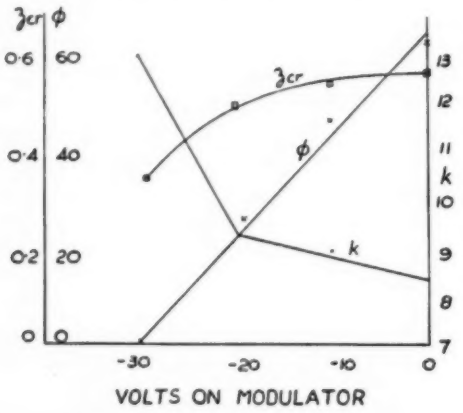


FIG. 4. Effect of modulation on crossover parameters.

with respect to the cathode. The family of curves obtained for negative voltages up to the blackout point follow the same exponential law as for $M=0$, but the value of "A" decreases while that of "k" increases in each case. A separate experiment in which the emitting area of the cathode was covered with a square gauze network showed clearly that one effect of the modulation was to reduce the radius of the emitting area of the cathode in proportion to the modulation swing. For example, if the system required -30 volts to black out the beam, then if -10 volts was applied to the modulating electrode, the emitting radius was restricted to 20/30 of its size when the modulator was at zero or cathode potential. Hence "a" in (8) decreases in proportion to the modulation, while "k" usually increases somewhat (Fig. 4).

The product "ak" which determines the angle is predominantly affected by the large variation in the emitting radius, and this is the controlling factor, so that the calculated angle φ varies almost linearly with V_m.

There is a further effect as regards the position of the crossover, z_{cr}, when the beam is modulated; it tends to move toward the cathode. This shift would be expected on purely qualitative grounds, since it is known that the equipotentials become more and more curved as the inter-electrode space becomes more negative. This, in turn produces larger deviations of the rays, i.e., shorter focal lengths; the crossover point will thus tend to move towards the cathode. With this type of system, the crossover remains on the side of the modulator remote from the cathode throughout the modulation range.

V. THE SIZE OF THE CROSSOVER

The intersection of the path of an electron which leaves the cathode at normal incidence and with zero velocity (at some height above its center) with the axis of the system defines the plane of the crossover for this velocity. All such electrons emerging at different heights will meet at this point under truly paraxial conditions. The upper limit to the size of the crossover will be defined by the group of electrons leaving the center of the cathode at an angle of 90°, and with the greatest velocity. The points at which these strike the plane of the crossover sets the limit to its boundary. The actual physical number involved here will be small since the emitter can be assumed to obey the Lambert cosine law. Further, for each direction of emission, the Maxwell distribution will determine the fraction of the total number with any assigned velocity; for example about 70 percent of the electrons are to be found with all velocities up to the most probable velocity of about 0.1 volt. It is therefore sufficient to examine the crossover sizes obtained for electron speeds of 0.1 volt as a working limit, since the area enclosed will embrace some 70 percent of all the electrons emitted by the source. Unfortunately, it was not possible to adhere to this value for the simple theory, since the initial electron velocity is fixed by the value of A, and this is about 0.5 v when the blackout voltage is 30 v. Even with this limitation, the theory is worth stating since it provides data for an upper limit to the size of the crossover. The size of the final image can be calculated from this, once the potential in the crossover plane is known.

In order to trace the rays which leave the center of the cathode at angles within the paraxial range, we use the original trajectory equation, but the constants will have different values because of the new boundary conditions. The initial linear field is again neglected here, but a rough calculation shows that by the time an electron has passed through it, its early parabolic path has now a slope within the paraxial range of the exponential focusing field, and the crossover height will not be affected by the height at which the ray starts. Since

$$r=e^{-(k/4)z}\{P\cos(k\sqrt{3}/4)z+Q\sin(k\sqrt{3}/4)z\}$$

from (3) the value of P is zero when $r=0$; for $z=0$.

$$\therefore\quad r=e^{-(k/4)z}\cdot Q\sin(k\sqrt{3}/4)z. \tag{9}$$

When $z=z_{cr}$, then $r=$radius of the paraxial crossover $=r_{cr}=\exp[-(k/4)z_{cr}]\cdot Q\cdot\sin(k\sqrt{3}/4)z_{cr}$. Now put $z_{cr}=8\pi/3k\sqrt{3}$

$$\therefore\quad r_{cr}=\exp[-(2\pi/3\sqrt{3})]\cdot Q\sin(2\pi/3). \tag{10}$$

This ray will pass through the periphery of the crossover. Before doing so, it may proceed to some point corresponding to z_M along the axis at which its height is maximum, while being focused toward its image on the axis. At this maximum height

$$(dr/dz)_{z=z_M}=0;$$

hence, from (9),

$$z_M=(4\pi/3k\sqrt{3})=z_{cr}/2. \tag{11}$$

This ray thus reaches its maximum height at a distance from the cathode equal to half the crossover distance. The height of the ray at this point is $\exp(-\pi/3\sqrt{3})\cdot Q\cdot\sqrt{3}/2$ and will obviously be governed, apart from its initial velocity "A," by the initial angle θ (assumed within the paraxial range) at which the electron leaves the cathode center, i.e., $(dr/dz)_{z=0}=\tan\theta$. Hence from (9),

$$Q=4\tan\theta/k\sqrt{3}; \tag{12}$$

$$\therefore\quad r=\exp[-(k/4)z]\cdot 4\tan\theta/k\sqrt{3}\cdot\sin(k\sqrt{3}/4)\cdot z. \tag{13}$$

This gives the height of the ray at any point in terms of the angle of ejection, and the geometry of the field

$$\therefore\quad r_{cr}=(2/k)\cdot\exp[-(2\pi/3\sqrt{3})]\cdot\tan\theta. \tag{14}$$

Hence, the paraxial crossover size is determined for electrons of initial velocity "A" by the ratio of $\tan\theta$ to "k." It might be supposed that there was no limit to the crossover size if θ took on the value of 90°. The theory would of course not hold for this case. Actually, the work with the differential analyzer described in the next section shows how quickly such rays are brought in by the field; the modulator hole does not appear to act in any way as a limiting stop. An examination of (14) indicates that the crossover size can be reduced only by a variation in geometry. In the standard C.R. tubes, "k" has already a value of about 10, and it is unlikely that it can be made much smaller by increasing the geometrical factor alone beyond this limit.

The aim in design should be to increase "k" to as large a value as possible, while at the same time insuring that the crossover is situated at a point in the field which has the lowest potential.

This should yield the minimum crossover image at the screen in theory; in practice, lens aberrations and space charge appear to be the controlling factors.

The maximum height of the trajectory was shown to be $\exp(-\pi/3\sqrt{3})\cdot Q\cdot\sqrt{3}/2$ and this occurs at a point approximately halfway to the crossover. All the rays leaving the center of the cathode will reach maximum height in the vicinity of this point; the peak height is found on substitution to be about 1.8 times the height of the crossover, independent of θ.

VI. TRAJECTORIES WITH THE DIFFERENTIAL ANALYZER

The crossover properties deduced from the simple theory can be helpful in design of systems, and the approximate form of the trajectory can also be traced out point by point for the paraxial case from (13).

In order to deal with electrons leaving the center of the cathode at angles *outside* the paraxial range, i.e., up to 90°, the differential analyzer was employed to solve the trajectory equation.

The general equation has been stated (1) as follows:

$$d^2r/dz^2+\{[1+(dr/dz)^2]/2(V+A)\}\cdot\partial V/\partial z\cdot dr/dz -\{[1+(dr/dz)^2]/2(V+A)\}\cdot\partial V/\partial r=0, \tag{15}$$

where V is the potential at any point on the axis and "A" is the initial electron velocity.

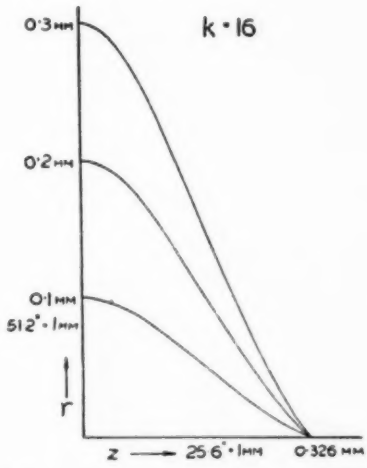


FIG. 5. Trajectories with differential analyzer.

This equation is not easily manageable unless paraxial conditions are imposed, e.g., $(dr/dz)\ll 0.1$, in which case it takes the form

$$d^2r/dz^2+\{V'/2(V+A)\}dr/dz +\{V''/4(V+A)\}\cdot r=0. \tag{16}$$

The differential analyzer solutions were, however, obtained from (16) which was taken in the simpler form of the two equations of motion, and before the time parameter "t" had been eliminated, i.e.,

$$m(d^2z/dt^2)=e(\partial V/dz), \tag{17}$$

$$m(d^2r/dt^2)=-er(\partial^2V/\partial z^2)/2. \tag{18}$$

If the electrons leave the cathode with initial velocity "A," then the results of the analyzer traces show that they follow approximately the same trajectory under the potential field $V=(Ae^{kz}-A)$, as those with zero velocity under a field $V=A(e^{kz}-e^{-kz})$. The value of "A" for cases of 30-v blackout is, as stated earlier, about 0.5 volt. This gives the diameter of the crossover calculated from (14) a practical significance. The analyzer traces were taken assuming the field to have the sinh-form all the way from the crossover to the cathode surface. The periphery of the crossover is then defined by the electrons which leave the center of the cathode at 90° and with the most probable velocity of 0.1 v.

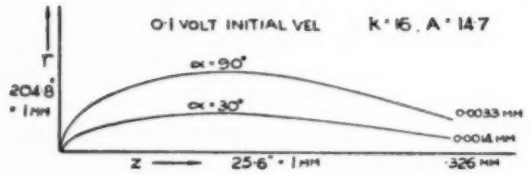


FIG. 6. Trajectories defining radius of crossover.

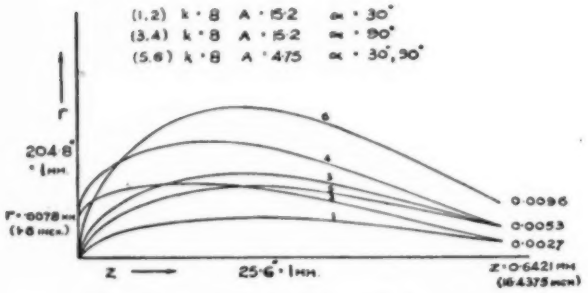


FIG. 7. Trajectories for 30° and 90° emission.

The electron distribution at the crossover assuming ideal focusing systems has already been treated by Langmuir[4] and will not be further discussed here.

In addition to the 90° trajectory, a further one of 30° was taken as a matter of interest, using the most probable velocity of 0.1 volt for the electrons leaving the center of the cathode. In order to find the crossover plane, it was necessary to trace electrons leaving the cathode normally at different heights above the axis. These are shown in Fig. 5 for a system with a value of $k=16$ and 5000 volts on the anode, and for an initial velocity of 0.1 volt. In order to keep within the paraxial region, the largest height of incidence used was 0.3 mm (emitting radius 0.5 mm). For the same system, the limiting value of the crossover radius is seen from Fig. 6 to be 0.0033 mm.

The size of the crossover will, for a given initial velocity, depend on the voltage applied to the system. If this is varied, the size is readily determined from the machine constants. For a 30-v blackout with this system and a 0.1 volt emission its value is 0.033 mm.

Figure 7 shows a series of trajectories for a system having $k=8$ and for various anode voltages, for angles

[4] D. B. Langmuir, Proc. I. R. E. **25**, 977 (1937).

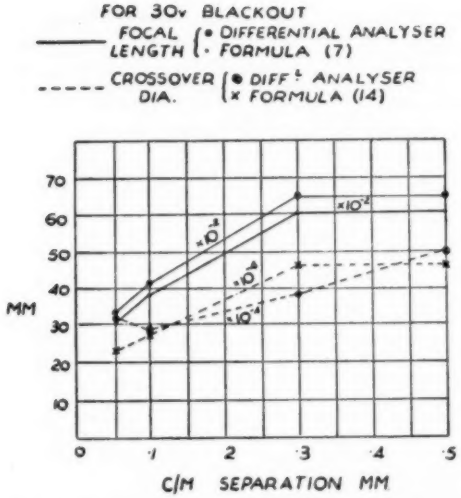


FIG. 8. Comparison between differential analyzer and simple theory.

of 30° and 90°. It will be observed from the pairs (1) and (2), and also (3) and (4), which relate to rays leaving the cathode surface with the same direction and velocity, but at different heights, that they are focused to the same point in the crossover plane, which is to be expected under these conditions. It should also be noticed that the trajectories from the center of the cathode reach a maximum height at a point roughly halfway to the crossover, and this is usually about twice the crossover height, as shown earlier.

The focal length of the lens formed by the immersion objective system is seen from Fig. 8 to increase with the C/M separation. The value calculated by formula (7) is shown to be a good approximation when compared with the traced value on the analyzer. The same figure shows how the crossover diameter varies as the cathode is moved away from the modulator keeping the blackout voltage constant at the value of 30 v. The approximation obtained from (14) is shown to be quite good, even though no rays are included which are inclined at an angle greater than 17°.

In conclusion, I wish to thank the Director of the National Physical Laboratory for his kindness in placing the facilities of the differential analyzer (housed in this University) at my disposal, and Messrs. Michel and Butler for helpful discussions during its operation.

Single Crystal Copper Surfaces*

T. N. Rhodin, Jr.
Institute for the Study of Metals, University of Chicago, Chicago, Illinois
(Received February 20, 1950)

Some surface structure characteristics of a metal single crystal are considered and the conclusions are applied to the preparation of single crystal plates of copper. A successful preparation is described as well as suitable methods for evaluating the effectiveness of the preparation. Using physical microadsorption isotherms of nitrogen at 78.1°K, electron diffraction, and x-ray diffraction methods, the conclusion is tentatively asserted that the surfaces are essentially planar on a molecular scale, and that the surface atoms are arranged in an approximately undistorted lattice.

I. INTRODUCTION

IT has become a question of great interest to ascertain the relationship between the physical and chemical properties of a metal interface and that particular crystallographic arrangement characteristic of the atoms in the interface area between the metal and its environment. The environment could be a fluid or solid such as another grain of the same metal. Although these remarks apply equally well to any such interface in general, let us for convenience limit consideration to a single crystal-gas interface. The surface of the metal crystal may, in a region where the macroscopic surface has a given orientation, approximate the ideal crystal plane of the corresponding orientation—that is to say, a surface formed from an infinite crystal lattice by removing all atoms whose centers lie on one side of the plane. This would give steps and risers of atomic dimensions only.[1] The surface may, on the other hand, have a hill and valley structure on a scale larger than atomic dimensions,[2,3] so that the planar arrangement of the microscopic surface may well differ from that of the macroscopic surface as illustrated in Fig. 1. Thermodynamic considerations prescribe that the surface structure with the lowest free energy will be the true one providing that the surface be in equilibrium with its vapor for a sufficiently long time. However, as is so often the case for the solid state, since the structure actually obtained after a finite time will depend on the rate of approach of the system to an equilibrium state, the conditions experimentally obtained at room temperature may differ considerably. Hence, although there is much evidence[4] for the unlikelihood of hill and valley structures and for the validity of the hypothesis that surfaces are atomically smooth if the temperature is sufficiently high, such observations do not necessarily apply to metal surfaces prepared at low temperatures. Yet the definition of surface structure is a critical prerequisite to surface studies and it is exceedingly difficult to do so, when room temperature surface polishing techniques tend to distort mechanically the structure of the surface layers.[5] On the other hand, if the surface distortion were minimized the surface orientation of sufficiently planar surfaces would be that of the matrix atoms. The preparation of copper surfaces of this type suitable for surface chemistry studies using a sensitive quartz vacuum microbalance was systematically investigated without resorting to high temperature equilibrium conditions. Weight increments caused by surface effects such as adsorption or oxidation are often small and, even for an exceedingly sensitive instrument, it is desirable for the samples to possess a high surface to volume ratio; in other words, the optimum geometry of the sample is that of a very thin plate. It is difficult enough to satisfy the conditions of planarity and undistorted surface structure under the most favorable conditions but special considerations are required when this third condition is added. Recently a systematic study of these considerations, applicable to thin single crystal plates of copper, was completed. A concise report of the results seems desirable.

II. PROCEDURE AND RESULTS

The original source of the crystal plates was a cylindrical single crystal of copper (6 in.×1¼ in.) grown by

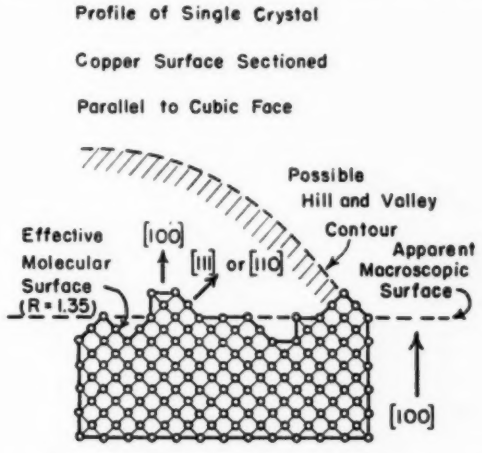


FIG. 1. Profile of single crystal copper surface sectioned parallel to cubic face (schematic). (Arrows indicate crystallographic directions normal to crystal face.)

* The research was supported in part by Army Air Forces Contract AF-33(038)-6534.

[1] S. T. Martin, Phys. Rev. **56**, 947 (1939).
[2] I. Langmuir, Phys. Rev. **22**, 357 (1923).
[3] L. Tonks, Phys. Rev. **38**, 1030 (1931).
[4] C. Herring and M. H. Nichols, Rev. Mod. Phys. **21**, 258 (1949).
[5] C. F. Elam, Trans. Faraday Soc. **32**, 1604 (1936); S. Dobinski, Nature **138**, 685 (1936).

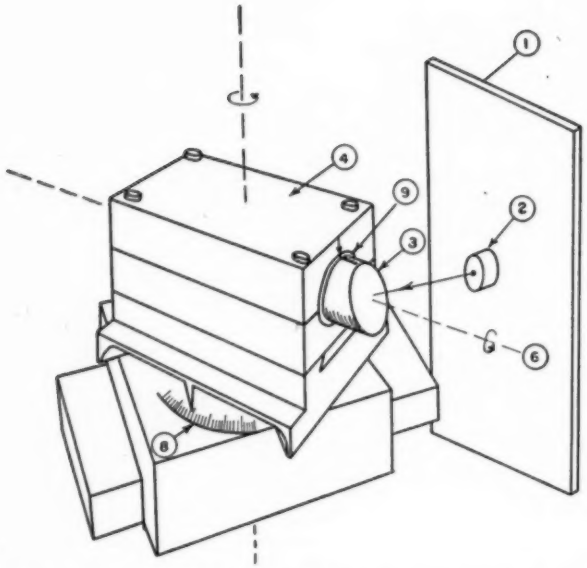


FIG. 2. Drawing of jig holder for orientation determination of cylindrical single crystals. (1) Film cassette; (2) x-ray pinhole; (3) crystal; (4) jig; (5) vertical axis rotation; (6) horizontal axis rotation; (7) cutting wheel; (8) graduated circle; (9) fiber spacer.

the standard Bridgman method. The orientation and symmetry of the crystal was determined by the Laue method, using back reflection x-rays. Testing the crystal at various spots by this method indicated an over-all variation of lattice orientation of two degrees or less. The copper was 99.995 percent pure with traces of iron, silicon, and aluminum. Sections ($\frac{3}{4}$ in$\times\frac{1}{2}$ in.) were cut from the crystal very slowly with an alundum cutting wheel at a proper angle so as to expose the macroscopic face along a predetermined low index plane. The proper transfer of the crystal from the Laue camera set-up to the cutting wheel was insured by the use of an (some of the techniques roughly resemble those used by A. T. Gwathmey and A. F. Benton, J. Phys. Chem. **44**, 35 (1940)) interchangeable jig arrangement schematically indicated in Figs. 2 and 3, respectively. The crystal was so arranged in the jig to permit one to orient the pole of the desired plane by simple rotations about two axes, one parallel to the cylinder axis and one normal to it, as indicated by items 5 and 6 in Fig. 2. Mounting the section in Lucite with one face exposed permitted one to polish it mechanically by the usual methods, finishing up with Buehler powder No. 3 on the polishing wheel. Upon reversal in the mount the second face was likewise polished until the thickness of the sample had been reduced to 0.040 in. This preliminary mechanical polish was necessary to insure uniform removal of material by the relatively prolonged electropolishing that followed. The plate was thereupon floated off the mount in benzene and the thickness reduced 75 percent to 0.010 in. by a carefully tedious process of electropolishing so as to insure removal of the distorted layers. The technique in this stage was a standard one except that the rate of removal of material was kept very small. A solution of orthophosphoric acid and ethyl alcohol was used in the usual proportions with an etching current of less than 0.01 ampere at about a potential of 2.0 volts for several periods of 20 minutes each until sufficient material had been removed. The long treatments tended to develop a wavy surface, but on such a large scale that it did not contribute to the molecular roughness. The final appearance was mirror bright and planar. No pitting nor selective attack of the surface was observed by microscopic observation at 1000 diameters.

The effectiveness of the preparation was tested in two ways. The planarity of the final surfaces on a molecular scale was determined from microadsorption isotherms of nitrogen at 78.1°K using a very sensitive microbalance technique described in detail elsewhere.[6] In brief, the amount of physically adsorbed gas as a function of pressure was measured and the weight of adsorbed gas corresponding to a monolayer coverage of the surface by a close packed layer of adsorbed gas molecules was calculated from the adsorption isotherms by the Brunauer-Emmett-Teller equation[7] and checked by the Harkins-Jura method.[8] Using the proper value for the cross-sectional area of an adsorbed nitrogen molecule a simple calculation yields the value of the effective surface area available for nitrogen molecule adsorption, that is to say, the molecular surface area. The ratio of this surface area to that calculated from the macroscopic dimensions of the sample gives a precise value for the roughness of the surface. For example, if (R) is the roughness factor, the percent roughness of the surface on

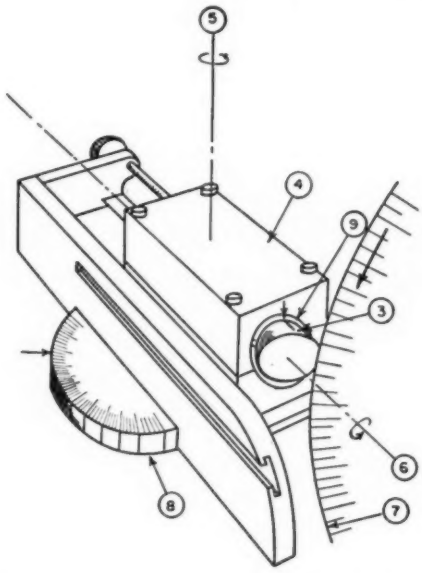


FIG. 3. Drawing of jig holder for sectioning of cylindrical single crystals (1) Film cassette; (2) x-ray pinhole; (3) crystal; (4) jig; (5) vertical axis rotation; (6) horizontal axis rotation; (7) cutting wheel; (8) graduated circle; (9) fiber spacer.

[6] T. N. Rhodin, J. Am. Chem. Soc. (to be published).
[7] S. Brunauer, *The Adsorption of Gases and Vapors*, Vol. I, *Physical Adsorption* (Princeton University Press, Princeton, New Jersey, 1942), p. 149 ff.
[8] W. Harkins, and G. Jura, J. Am. Chem. Soc. **66**, 1366 (1944).

a molecular scale may be expressed as $100(R\text{-}1)$. Since it is reasonable to assume for a metal surface that irregularities will tend to be more or less spread out over the whole surface, surfaces with roughness factors near unity will be essentially planar. Such was the case for the surfaces prepared as described, where most of the R values ranged from 1.2 to 1.5.

The perfection of the surface structure was examined directly by both glancing angle electron diffraction and transmission and back-reflection of x-rays. No definable diffraction pattern was obtainable from the electron diffraction examinations, although the surface had been carefully cleaned by distilled water and alcohol rinses followed by a 400 C. hydrogen reduction and outgassing pre-treatment. Back-reflection x-ray patterns indicated a maximum lattice distortion of two degrees averaged over approximately the outer 5000 layers of copper atoms. In order to define better the structure of the surface layers, transmission x-ray examinations were made of a small section in the center of a crystal plate which was systematically reduced in thickness by successive electropolishing treatments. The ultimate thickness examined appeared to be less than 10^{-4} cm with no indication that the observed distortion was not more or less characteristic of the whole thickness of the plate.

III. DISCUSSION

If one assumes that the surfaces used for electron diffraction examination were adequately clean, the inability to observe coherent reflection patterns may just indicate the absence of suitable projections on the surface for the electrons to penetrate without being totally scattered. This might be conceivably true for surfaces approximately planar on a molecular scale. In addition, x-ray reflection and transmission patterns, indicating a maximum variation in lattice perfection of two degrees, may be considered to indicate a relatively undistorted surface layer providing one accepts the admittedly inconclusive indication that the observed small distortion is not intensively concentrated in the surface layers. It cannot be denied that the results do not preclude the possibility that a highly distorted outermost layer, too thin to contribute to the x-ray pattern, may exist. In addition, well-defined reflection patterns have been obtained for glancing angle electron reflection from planar surfaces of mica and from thin films of dielectrics formed by condensation on to mica substrates.[9] However, the existence of this effect for unetched single crystal metal surfaces has yet to be established and the present inability to observe it remains an open question.

In the final analysis the fact that atoms are arranged in a particular crystallographic plane on the surface may be best indicated by some physical chemical effect on a molecular scale, the magnitude of which can be measured and the extent of which is sensitively dependent on the particular crystallographic face involved. Two such effects, namely, low temperature oxidation, and low temperature physical adsorption of gases have been investigated in this laboratory. In both cases the magnitude of the energy terms involved, e.g., the activation energy for oxidation and the heat of physical adsorption have been observed to depend on the crystallographic direction of the macroscopic surface as if most of the surface atoms were arranged in a characteristic well-ordered surface lattice.[10] Since both phenomena occur by a molecular mechanism, there is reason to believe that there exists a precise and reproducible dependence on the surface structure. Although a detailed discussion of these effects is out of place here, the fact that they have been observed is cited as evidence for the assumption that the surface structures of the samples prepared, as previously described, correspond to a considerable extent to that of a fairly well-ordered surface lattice.

IV. CONCLUSIONS

It is believed that thin relatively undistorted single crystal plates of copper can be prepared from large single crystals. A suitable process of mounting and electropolishing is described, as well as a sample holder to facilitate orientation determination and sectioning. There is evidence to believe that the surfaces are essentially planar and that the surface atoms exist in an orientation corresponding to a considerable extent to that characteristic of parallel crystallographic planes in the matrix of the crystal.

V. ACKNOWLEDGMENTS

The author is grateful to K. Ikeuye, J. Hess, and L. Schulz for discussions and assistance. The painstaking technique required for the preparation of the samples was furnished by the patient effort of J. Cerny and is also gratefully acknowledged.

[9] L. G. Schulz, J. Chem. Phys. **18**, 996 (1950).
[10] T. N. Rhodin, J. Am. Chem. Soc. (to be published).

Theory of the Parallel Plane Diode

A. H. Taub and Nelson Wax
University of Illinois, Urbana, Illinois
(Received March 10, 1950)

Transient and steady state solutions of the parallel plane diode are obtained. The limitations on a single valued velocity theory are investigated. These general results are applied to the space charge limited diode with the a.c. components small compared to the d.c. Power consumption is discussed and it is shown that hitherto neglected terms are of importance.

I. INTRODUCTION

THE parallel plane diode has been the subject of a large number of investigations, both theoretical and experimental.[1–6] A good deal of attention has been centered on problems concerning the behavior of the plane diode when both space charge and time-varying fields are present. These latter studies usually[4–6] have employed perturbation techniques: the time-varying quantities were assumed to be first-order corrections on the static quantities.

It is the purpose of this paper to discuss the theory of the plane diode when a time dependent current density flows between the cathode and plate. The following assumptions are made throughout: (1) The diode consists of an infinite plane cathode parallel to an infinite plane anode (edge effects are ignored). (2) The magnetic effects of the moving electrons are neglected.

Under these assumptions the fundamental equations are reduced to dimensionless form in II. The transient behavior of the diode is treated in III, and it is shown that the differential equations of the system may be solved by quadratures if the initial position, x^0, and velocity, v^0, of every electron is known at the time, $t=0$, when the current starts to flow. The steady state behavior is discussed in IV; the equations may be integrated directly again when ξ, the time of emission of an electron from the cathode, and t, the time, are chosen as independent variables. The necessary and sufficient condition that the electron trajectories shall not intersect is obtained as an immediate consequence of this formulation. The electron transit time and the plate voltage are given as functions of ξ and t, in V and VI, respectively. The treatment in these, and in subsequent sections, is restricted to a small signal analysis of the space charge limited diode. Fourier expansions for all the quantities which were discussed in earlier sections are given in VII, and these expansions are used in VIII to obtain an expression for the average power consumption of the diode per unit area. It is shown that the notion of complex impedance as used by previous workers is insufficient alone to account for all the power consumed by the diode.

All dimensional equations are in m.k.s. units.

II. THE FUNDAMENTAL EQUATIONS

The plane diode is described by the following three differential equations. The equation of motion of an electron in the diode is

$$(d^2x'/dt'^2)=(e/m)E', \tag{2.1}$$

where x' is the distance of the electron from the cathode, t' the time, E' the electric field intensity, and e/m the charge to mass ratio of an electron.

The equation of conservation of charge is

$$\partial(\rho'v')/\partial x'+\partial\rho'/\partial t'=0 \tag{2.2}$$

with ρ' the charge density and v' the velocity of the electrons associated with ρ'. $v'=dx'/dt'$.

Poisson's equation takes the form

$$\partial E'/\partial x'=\rho'/\epsilon_0. \tag{2.3}$$

ϵ_0 is the dielectric constant of free space $=1/36\pi\times10^{-9}$ farad per meter.

The total current density, I, is defined by

$$I=\rho'v'+\epsilon_0(\partial E'/\partial t'). \tag{2.4}$$

The first term in the right-hand member of Eq. (2.4) is the conduction current density and the second is the displacement current density. It is of particular note that I is a function of t' alone, and not of x'. This result may be obtained by substituting ρ', from Eq. (2.3), into Eq. (2.2) and using the definition for I (see Eq. 3.14).

If I is a periodic function of t', with period τ, then one can introduce the dimensionless set of variables

$$x=x'/a, \tag{2.5}$$

and

$$t=t'/\tau, \tag{2.6}$$

[1] W. E. Benham, Phil. Mag. **5**, 641 (1928); **11**, 457 (1931).
[2] J. Müller, Hochfrequenztech. u. Elektroakustik **41**, 156 (1933); **43**, 195 (1934).
[3] F. B. Llewellyn, *Electron Inertia Effects* (Cambridge University Press, London, England, 1943), contains a bibliography.
[4] H. W. König, Hochfrequenztech. u. Elektroakustik **62**, 76 (1943).
[5] F. Gray of the Bell Telephone Laboratories Inc. has obtained the general solution of the diode equations by a different method in an unpublished memorandum dated 11/11/41. This material was presented as an invited paper "The general solution for an electron stream in a parallel plane diode" at the 266th meeting of the Am. Phys. Soc. June 15–16, 1945. Gray did not interpret his solutions in the fashion presented here. König (reference 4) has used an approach which is similar to, but more restricted than, Gray's in discussing the system of differential equations (2.11)–(2.13).
[6] H. W. König, *Laufzeittheorie der Elektronenrohren* (Verlag. Julius Springer, Vienna, 1948), Vol. 1.

where a is the distance between the cathode and plate of the diode.

The total current density may now be written as

$$I=I_0[1+\mu f(t)]=I_0 h(t). \tag{2.7}$$

I_0 and μ are constants, and $f(t)$ is an arbitrary periodic function of period one in t (τ in t'), and of maximum absolute value one. I_0 is thus the direct current density, and μI_0 is the maximum amplitude of the time dependent current density.

The dimensionless quantities v, E, and ρ are defined by

$$v=(\tau/a)v', \tag{2.8}$$

$$E=(\epsilon_0/\tau I_0)E' \tag{2.9}$$

and

$$\rho=(a/\tau I_0)\rho', \tag{2.10}$$

respectively. Equations (2.1) to (2.3) may be written in dimensionless form as

$$d^2x/dt^2=(6/\kappa^3)E, \tag{2.11}$$

$$\partial(\rho v)/\partial x+\partial\rho/\partial t=0, \tag{2.12}$$

and

$$\partial E/\partial x=\rho, \tag{2.13}$$

respectively. $\kappa^3=(6am\epsilon_0/eI_0\tau^3)$ is a dimensionless constant.

III. THE TRANSIENT CASE

Let the position x^0 and the velocity v^0 of every electron be known, in the region between cathode and plate, at the time $t=0$, the instant at which the known current I starts to flow. At time $t=t_1$ the electron which emerged from the cathode at $t=0$ will be collected at the plate. It is the object of this section to study the behavior of those electrons which were in the anode-cathode region, at $t=0$, in the time interval $0\leq t\leq t_1$.

If x^0 is known, then ρ^0, the density distribution of the electrons at $t=0$, is known. Note that $v^0=v^0(x^0)$ and $\rho^0=\rho^0(x^0)$. The subsequent position of any electron may be represented by

$$x=x(x^0, t), \tag{3.1}$$

where x^0 is its position at $t=0$.

Equation (3.1) together with the equation

$$t=t \tag{3.2}$$

may be considered as a transformation of coordinates from the variables x, t to the variables x^0, t. The velocity of every electron may be considered to be a function of both x^0 and t. Indeed Eq. (3.1) is the solution of

$$dx/dt=v(x^0, t). \tag{3.3}$$

Furthermore, if one fixes one's attention on any individual electron, then one has, using Eqs. (3.1) and (3.3), that

$$dx/dt=\partial x/\partial t=v(x^0, t), \tag{3.4}$$

since x^0 is fixed for each electron.

The matrix, M, of the Jacobian of the transformation defined by Eqs. (3.1) and (3.2) is

$$M=\begin{Vmatrix}\partial x/\partial x^0 & \partial x/\partial t\\ 0 & 1\end{Vmatrix}=\begin{Vmatrix}\partial x/\partial x^0 & v\\ 0 & 1\end{Vmatrix}. \tag{3.5}$$

The matrix, M', of the Jacobian of the inverse transformation (i.e., the transformation which gives x^0 and t in terms of x and t) is

$$M'=\begin{Vmatrix}\partial x^0/\partial x & \partial x^0/\partial t\\ 0 & 1\end{Vmatrix}=\begin{Vmatrix}\dfrac{1}{\partial x/\partial x^0} & -\dfrac{v}{\partial x/\partial x^0}\\ 0 & 1\end{Vmatrix}. \tag{3.6}$$

Thus, for any function $F(x, t)$ one has

$$\frac{\partial F}{\partial x}=\frac{\partial F}{\partial x^0}\frac{\partial x^0}{\partial x}=\frac{\partial F}{\partial x^0}\frac{1}{\partial x/\partial x^0} \tag{3.7}$$

and

$$\frac{\partial F}{\partial t}=-\frac{v}{\partial x/\partial x^0}\frac{\partial F}{\partial x^0}+\frac{dF}{dt}. \tag{3.8}$$

The notation dF/dt used in Eq. (3.8) is to denote the derivative of F with respect to t keeping x^0 constant, i.e., following a particular electron along its path. Furthermore,

$$\frac{dF}{dt}=\frac{\partial F}{\partial t}+v\frac{\partial F}{\partial x}, \tag{3.9}$$

when the variables x and t are used.

The equation of the conservation of charge is equivalent to the statement,

$$\frac{\partial x}{\partial x^0}=\frac{\rho^0(x^0)}{\rho}, \tag{3.10}$$

if x is a unique function of x^0 and conversely that x^0 is a unique function of x. One assumes, in writing Eq. (3.10), that M, and therefore M', are non-singular. This implies that electron trajectories cannot cross or touch each other.

The Poisson equation now takes the form

$$\partial E/\partial x^0=\rho^0(x^0), \tag{3.11}$$

and the equation of motion remains unchanged

$$d^2x/dt^2=(6/\kappa^3)E. \tag{3.12}$$

The general solution of Eq. (3.11) is

$$E=\int_0^{x^0}\rho^0(\eta)d\eta+g(t)+E_0, \tag{3.13}$$

where E_0 is the electric field intensity at $x^0=0$ and $g(t)$ is an arbitrary function of time. The particular solution

for the diode may be obtained by noting that

$$\frac{dE}{dt}=v\frac{\partial E}{\partial x}+\frac{\partial E}{\partial t}=\rho v+\frac{\partial E}{\partial t}=\frac{I}{I_0}=\frac{dg}{dt}=h(t). \quad (3.14)$$

$h(t)$ is a function of t alone, as was mentioned previously. Thus

$$E=\int_0^t h(t')dt'+\int_0^{x^0}\rho^0(\eta)d\eta+E_0 \quad (3.15)$$

for the diode. Equation (3.12) may be integrated directly. One has

$$\frac{\kappa^3}{6}v=\int_0^t\int_0^{t'}h(t'')dt''dt'+t\left(\int_0^{x^0}\rho^0(\eta)d\eta+E_0\right)+\frac{\kappa^3}{6}v^0(x^0), \quad (3.16)$$

where $v^0(x^0)=v(x^0, 0)$, and

$$\frac{\kappa^3}{6}x=\int_0^t\int_0^{t'}\int_0^{t''}h(t''')dt'''dt''dt'+\frac{t^2}{2}\left(\int_0^{x^0}\rho^0(\eta)d\eta+E_0\right)+[tv^0(x^0)+x^0]\frac{\kappa^3}{6}. \quad (3.17)$$

ρ may be obtained from Eqs. (3.17) and (3.10) if desired.

The above solution was obtained under the assumption that M is non-singular. However those electrons which were at the cathode, $x^0=0$, at $t=0$ have orbits which differ; $x^0=0$ is a singular point, therefore, and the solution is inapplicable for electrons emitted when $t>0$. Equations (3.15) to (3.17) are valid only when $0<x^0\leq 1$. The orbit of the electron emitted at $t=0$ can be obtained by a limiting operation:

$$x(t)=\lim_{x^0\to 0}x(x^0, t). \quad (3.18)$$

IV. THE STEADY STATE CASE

Those electrons which are emitted from the cathode for $t>0$ were not under consideration in the previous section but will be studied now. The transformation given by Eqs. (3.1) and (3.2) proved to be singular at the cathode, and it is necessary to use some other variable rather than the initial positions of the electrons, to distinguish them. If the dimensionless variable ξ is chosen, corresponding to the time when an electron is emitted from the cathode, then any electron's orbit is described by

$$x=x(\xi, t). \quad (4.1)$$

One uses the equation

$$t=t \quad (4.2)$$

again, and the pair of equations, (4.1) and (4.2), yield an analogous transformation to Eqs. (3.1) and (3.2). All the preceding formulas apply when x^0 is replaced by ξ.

Equation (3.15) becomes

$$E(\xi, t)=\int_\xi^t h(t')dt'+E_0(\xi), \quad (4.3)$$

where $E_0(\xi)$ is the electric field intensity at the cathode ($t=\xi$). Similarly

$$\frac{\kappa^3}{6}v(\xi, t)=\int_\xi^t\int_\xi^{t'}h(t'')dt''dt'+(t-\xi)E_0(\xi)+v_0(\xi)\kappa^3/6=\int_\xi^t(t-t')h(t')dt'+(t-\xi)E_0(\xi)+v_0(\xi)\kappa^3/6 \quad (4.4)$$

is the solution for $v(\xi, t)$ when $v_0(\xi)$ is the velocity with which electrons are emitted from the cathode.

The orbit is given by

$$\frac{\kappa^3}{6}x(\xi, t)=\int_\xi^t\int_\xi^{t'}\int_\xi^{t''}h(t''')dt'''dt''dt'+\frac{(t-\xi)^2}{2}E_0(\xi)+(t-\xi)v_0(\xi)\kappa^3/6=\int_\xi^t\frac{(t-t')^2}{2}h(t')dt'+\frac{(t-\xi)^2}{2}E_0(\xi)+(t-\xi)v_0(\xi)\kappa^3/6 \quad (4.5)$$

since $x(\xi, \xi)=0$; namely, this corresponds to the cathode.

The matrix of the transformation given by Eqs. (4.1) and (4.2) is non-singular if and only if

$$\partial x/\partial\xi\neq 0. \quad (4.6)$$

Equations (4.6) and (4.5) lead to the condition† that

$$-\frac{(t-\xi)^2}{2}\left[h(\xi)-\frac{dE_0(\xi)}{d\xi}\right]+(t-\xi)\left[\frac{\kappa^3}{6}\frac{dv_0(\xi)}{d\xi}-E_0(\xi)\right]-\frac{\kappa^3}{6}v_0(\xi)\neq 0 \quad (4.7)$$

for all t and ξ with $t\neq\xi$, since at the cathode all orbits intersect. If $E_0(\xi)$ and $v_0(\xi)$ are constant in time then (4.7) becomes

$$\frac{\kappa^3}{6}v_0(\xi)+\frac{(t-\xi)^2}{2}h(\xi)+(t-\xi)E_0(\xi)\neq 0. \quad (4.8)$$

† J. Müller (reference 2) has obtained this as a sufficient condition by different methods.

The subsequent discussion will apply only to the space charge limited diode, for which

$$E_0(\xi)=v_0(\xi)\equiv 0. \tag{4.9}$$

One has then that orbits will not intersect away from the cathode ($t\neq\xi$) if and only if

$$I(\xi)\neq 0. \tag{4.10}$$

Whenever the current density is represented by Eq. (2.7) then it follows from (4.10) that

$$\mu<1 \tag{4.11}$$

is the necessary and sufficient condition that any two electron orbits shall not have a point in common in the space between cathode and plate.

The inequality (4.11) will be assumed throughout the remainder of this paper.

The space charge density distribution for the space charge limited diode is, from Eq. (2.13),

$$\rho(\xi,t)=\frac{\partial E}{\partial x}=\frac{\partial E}{\partial \xi}\Big/\frac{\partial x}{\partial \xi}=-h(\xi)\Big/\frac{\partial x}{\partial \xi}, \tag{4.12}$$

or

$$\rho(\xi,t)=\frac{2\kappa^3}{6(t-\xi)^2}. \tag{4.13}$$

Whenever ξ is known as a function of x and t, then ρ can be given in terms of these variables. The dependence of ξ on x and t is obtained from the inversion of Eq. (4.5), an operation which is feasible, generally, by using graphical or numerical methods.

V. TRANSIT TIME

The time which is required for an electron to travel the distance between cathode and plate is defined as the transit time. Let ξ_1 be the time at which an electron emerges from the cathode, and t the time at which it arrives at the plate; then

$$\frac{\kappa^3}{6}=\int_{\xi_1}^{t}\frac{(t-t')^2}{2}h(t')dt', \tag{5.1}$$

since $x=1$ is the coordinate of the plate, by Eq. (2.5). Equation (5.1) determines ξ_1, when $h(t)$ is known.

By definition the transit time T is

$$T=t-\xi_1. \tag{5.2}$$

If $h(t)$ is of the form given by Eq. (2.7) then Eq. (5.1) becomes

$$\kappa^3=(t-\xi_1)^3+3\mu\int_{\xi_1}^{t}(t-t')^2f(t')dt'. \tag{5.3}$$

Furthermore, if one considers the static diode, in which the current is constant, then Eq. (5.3) reduces to the Child-Langmuir solution

$$T=\kappa. \tag{5.4}$$

It is necessary to invert Eq. (5.1) for ξ_1 in order that T be known explicitly. One important property of T may be deduced without performing the inversion: if $h(t)$ is periodic then T will be periodic and of the same period.

The dimensionless period is unity and therefore

$$\begin{aligned}\frac{\kappa^3}{6}&=\int_{\xi_1(t+1)}^{t+1}\frac{(1+t-t')^2}{2}h(t')dt'\\&=\int_{\xi_1(t+1)-1}^{t}\frac{(t-t'')^2}{2}h(t''+1)dt''\\&=\int_{\xi_1(t+1)-1}^{t}\frac{(t-t'')^2}{2}h(t'')dt''\\&=\int_{\xi_1(t)}^{t}\frac{(t-t')^2}{2}h(t')dt'.\end{aligned} \tag{5.5}$$

It follows that

$$T(t+1)=t+1-\xi_1(t+1)=t-\xi_1(t)=T(t). \tag{5.6}$$

ξ_1, and hence T, may be obtained explicitly when μ is small. One expands ξ_1 in a MacLaurin series

$$\xi_1=\xi_0+\mu\left(\frac{d\xi_1}{d\mu}\right)_0+\frac{\mu^2}{2!}\left(\frac{d^2\xi_1}{d\mu^2}\right)_0+\cdots, \tag{5.7}$$

where the derivatives are evaluated at $\mu=0$, as is denoted by the subscript zero.

$$\xi_0=t-\kappa. \tag{5.8}$$

All the derivatives may be obtained from Eq. (5.3). For example, the first derivative is

$$\frac{d\xi_1}{d\mu}=\frac{\int_{\xi_1}^{t}(t-t')^2f(t')dt'}{(t-\xi_1)^2[1+\mu f(\xi_1)]} \tag{5.9}$$

with

$$\left(\frac{d\xi_1}{d\mu}\right)_0=\frac{\int_{\xi_0}^{t}(t-t')^2f(t')dt'}{\kappa^2}. \tag{5.10}$$

VI. THE PLATE VOLTAGE

At any fixed instant of time an electric field intensity exists at every point between cathode and plate. Minus one times the integral of this intensity with respect to distance, between two points, is defined as the voltage difference between these points for the fixed time under

consideration. Thus the plate voltage is, by definition,

$$V'(t')=-\int_0^a E'(x', t')dx'$$

$$=-\frac{a\tau I_0}{\epsilon_0}\int_0^1 E(x, t)dx=\frac{a\tau I_0}{\epsilon_0}V(t), \quad (6.1)$$

where the integrals are evaluated for the fixed time, t.

The dimensionless plate voltage, $V(t)$, may be written as

$$V(t)=\frac{3}{\kappa^3}\int_{\xi_1(t)}^t E(\xi, t)(t-\xi)^2h(\xi)d\xi. \quad (6.2)$$

As before, $V(t)$ will be periodic with the same period as $h(t)$, if $h(t)$ is periodic. One has

$$V(t+1)=\frac{3}{\kappa^3}\int_{\xi_1(t+1)}^{t+1} E(\xi, t+1)(t+1-\xi)^2h(\xi)d\xi$$

$$=\frac{3}{\kappa^3}\int_{\xi_1(t+1)-1}^t E(\eta+1, t+1)$$

$$\times(t-\eta)^2h(\eta+1)d\eta=V(t). \quad (6.3)$$

Use has been made of the periodicity of the transit time, and of the equation

$$E(\xi+1, t+1)=E(\xi, t), \quad (6.4)$$

which can be verified readily by referring to Eq. (4.3) and noting the $E_0(\xi)=0$.

Equation (6.2) is an expression for the voltage in terms of the known current density. The inverse relation is desired in many problems, that is, the dependence of the current on the voltage. Unfortunately the inversion of Eq. (6.2) involves the inversion of Eq. (5.1), which presents considerable analytical difficulty if the current density is an arbitrary function of time. These difficulties may disappear under some circumstances, but this problem will not be treated. The current density is assumed to be known.

The MacLaurin expansion for $V(t)$ is

$$V(t)=V_0+\mu\left(\frac{dV}{d\mu}\right)_0+\frac{\mu^2}{2!}\left(\frac{d^2V}{d\mu^2}\right)_0+\cdots \quad (6.5)$$

with

$$V_0=\frac{3}{\kappa^3}\int_{\xi_0}^t(t-\xi)^2d\xi=\frac{3\kappa}{4}. \quad (6.6)$$

The subscript notation is the one employed before, in Section V. The first derivative can be calculated by using Eqs. (6.2) and (5.9). One obtains

$$\frac{\kappa^3}{3}\frac{dV}{d\mu}=\int_{\xi_1}^t(t-\xi)^2\left[\frac{dE}{d\mu}h(\xi)+E\frac{dh}{d\mu}\right]d\xi$$

$$-E(\xi, t)\int_{\xi_1}^t(t-\xi)^2\frac{dh}{d\mu}d\xi. \quad (6.7)$$

This may be simplified by integrating by parts twice. The result is

$$\frac{\kappa^3}{6}\frac{dV}{d\mu}=\int_{\xi_1}^t[1+\mu f(\xi)]\int_\xi^t(t-t')$$

$$\times\int_{t'}^t f(t'')dt''dt'd\xi, \quad (6.8)$$

with

$$\frac{\kappa^3}{6}\left(\frac{dV}{d\mu}\right)_0=\int_{\xi_0}^t\int_\xi^t\int_{t'}^t(t-t')f(t'')dt''dt'd\xi$$

$$=\int_{\xi_0}^t\int_\xi^t(t-t')\mathcal{E}(t', t)dt'd\xi, \quad (6.9)$$

where

$$\mathcal{E}(\xi, t)=\int_\xi^t f(t')dt'.$$

The second derivative is given ultimately by

$$\frac{\kappa^3}{6}\frac{d^2V}{d\mu^2}=\frac{2}{(t-\xi_1)^2}\left[\int_{\xi_1}^t(t-t')\int_{t'}^t f(t'')dt''\right]^2$$

$$-\int_{\xi_1}^t(t-\xi)\mathcal{E}^2(\xi, t)d\xi, \quad (6.10)$$

with

$$\frac{\kappa^3}{6}\left(\frac{d^2V}{d\mu^2}\right)_0=\frac{2}{\kappa^2}\left[\int_{\xi_0}^t(t-\xi)\mathcal{E}(\xi, t)d\xi\right]^2$$

$$-\int_{\xi_0}^t(t-\xi)\mathcal{E}^2(\xi, t)d\xi. \quad (6.11)$$

Higher derivatives may be obtained in a similar fashion.

VII. FOURIER EXPANSIONS

Fourier expansions are given in this section for the electric field intensity, the position of an electron, various derivatives of the voltage, and for an auxiliary function of some usefulness in later discussions, when the Fourier series for the current density is known.

It should be remarked first that there is no loss in generality in taking $f(t)$ a periodic function of period one and zero mean. One may write then

$$f(t)=\sum_{-\infty}^{\infty}f_ne^{i\omega_nt}=\sum_{n=1}^{\infty}a_n\cos\omega_nt+b_n\sin\omega_nt \quad (7.1)$$

with $\omega_n=2\pi n$ and $f_0=0$. Since $f(t)$ is a real function of the real variable t, then f_n and f_{-n} are conjugate complex quantities, and the a_n and b_n are real.

Equation (4.3) becomes

$$\mathcal{E}(\xi, t)=\sum_{-\infty}^{\infty}\frac{f_n}{i\omega_n}(e^{i\omega_nt}-e^{i\omega_n\xi}). \quad (7.2)$$

The position x is given by

$$\frac{\kappa^3}{6}x(\xi,t)=\frac{(t-\xi)^3}{6}+\mu\sum_{-\infty}^{\infty}\frac{if_n}{\omega_n}e^{i\omega_n t}\left[\frac{(t-\xi)^2}{2}e^{-i\omega_n(t-\xi)}-\frac{i}{\omega_n}(t-\xi)e^{-i\omega_n(t-\xi)}+\frac{1}{\omega_n^2}(1-e^{-i\omega_n(t-\xi)})\right]. \quad (7.3)$$

It is convenient to introduce the function $G(\xi, t)$, defined by

$$G(\xi,t)=\int_\xi^t\int_{t'}^t(t-t')f(t'')dt''dt'=\int_\xi^t(t-t')\mathcal{E}(t',t)dt'. \quad (7.4)$$

Its Fourier series is

$$G(\xi,t)=\sum_{-\infty}^{\infty}-\frac{if_n}{\omega_n}e^{i\omega_n t}\left[\frac{(t-\xi)^2}{2}-\frac{i(t-\xi)}{\omega_n}\times e^{-i\omega_n(t-\xi)}+\frac{1}{\omega_n^2}(1-e^{-i\omega_n(t-\xi)})\right] \quad (7.5)$$

with

$$G(\xi_0,t)=\sum_{-\infty}^{\infty}B_n e^{i\omega_n t}, \quad (7.6)$$

$$B_n=-\frac{if_n}{\omega_n}\left[\frac{\kappa^2}{2}-\frac{i\kappa}{\omega_n}e^{-i\omega_n\kappa}+\frac{1}{\omega_n^2}(1-e^{-i\omega_n\kappa})\right]. \quad (7.7)$$

It follows from Eqs. (7.4) and (6.9) that

$$\frac{\kappa^3}{6}\left(\frac{dV}{d\mu}\right)_0=\int_{\xi_0}^t G(\xi,t)d\xi=\sum_{-\infty}^{\infty}-\frac{if_n}{\omega_n}e^{i\omega_n t}\left[\frac{\kappa^3}{6}+\frac{\kappa}{\omega_n^2}(1+e^{-i\omega_n\kappa})+\frac{2i}{\omega_n^3}(1-e^{-i\omega_n\kappa})\right]. \quad (7.8)$$

The Fourier series for the second derivative will not be given in detail. As can be seen from Eq. (6.11) and the definition for $G(\xi, t)$ given above, the second derivative contains the square of the series for $G(\xi, t)$ and an additional series. This last expansion may be written as

$$\int_{\xi_0}^t(t-\xi)\mathcal{E}^2(\xi,t)d\xi=\sum_{-\infty}^{\infty}G_n e^{i\omega_n t}, \quad (7.9)$$

where

$$G_n=\sum_{-\infty}^{\infty}-\frac{f_r f_{n-r}}{\omega_n\omega_{n-r}}\left\{\frac{\kappa^2}{2}-i\kappa\left[\frac{e^{-i\omega_r\kappa}}{\omega_r}+\frac{e^{-i\omega_{n-r}\kappa}}{\omega_{n-r}}-\frac{e^{-i\omega_n\kappa}}{\omega_n}\right]-\left[\frac{e^{-i\omega_r\kappa}-1}{\omega_r^2}+\frac{e^{-i\omega_{n-r}\kappa}-1}{\omega_{n-r}^2}-\frac{e^{-i\omega_n\kappa}-1}{\omega_n^2}\right]\right\}. \quad (7.10)$$

VIII. AVERAGE POWER

The power exchange between the d.c. and the a.c. components is of considerable interest. Earlier workers have shown that a negative resistance may occur if the diode is operating in a proper regime,[2-4,6] and a few self-excited oscillators have been built which make use of this phenomenon.[2-3] This section will be concerned with these questions.

The general expressions given above enable one to compute power for all $0\leq\mu<1$. As has been mentioned before, the analytic treatment is difficult when μ is not small. Attention will be restricted, therefore, to a small signal analysis.

The instantaneous power per unit area P' is defined as

$$P'=V'(t')I(t') \quad (8.1)$$

or

$$P'=\frac{a\tau I_0^2}{\epsilon_0}V(t)[1+\mu f(t)]=\frac{a\tau I_0^2}{\epsilon_0}P(t). \quad (8.2)$$

The average power per unit area, $\langle P'\rangle$, is the time average of P' over a cycle, and is given by

$$\langle P'\rangle=\frac{1}{\tau}\int_0^\tau P'(t')dt'=\frac{a\tau I_0^2}{\epsilon_0}\int_0^1 P(t)dt=\frac{a\tau I_0^2}{\epsilon_0}\langle P\rangle. \quad (8.3)$$

The MacLaurin series expansion for $\langle P\rangle$ is

$$\langle P\rangle=\langle P\rangle_0+\mu\left(\frac{d\langle P\rangle}{d\mu}\right)_0+\frac{\mu^2}{2!}\left(\frac{d^2\langle P\rangle}{d\mu^2}\right)_0+\cdots, \quad (8.4)$$

with the usual subscript notation.

The use of Eqs. (8.3), (6.9), and (6.11) yields the set of equations

$$\langle P\rangle=\frac{3\kappa}{4}, \quad (8.5)$$

$$\left(\frac{d\langle P\rangle}{d\mu}\right)_0=\int_0^1\left(\frac{dV}{d\mu}\right)_0dt+\int_0^1 V_0 f(t)dt=0, \quad (8.6)$$

$$\left(\frac{d^2\langle P\rangle}{d\mu^2}\right)_0=\int_0^1\left(\frac{d^2V}{d\mu^2}\right)_0dt+2\int_0^1\left(\frac{dV}{d\mu}\right)_0 f(t)dt. \quad (8.7)$$

Equation (8.6) states that the first-order correction to the d.c. power is zero. The second-order correction con-

tains two integrals which are, as will be shown, of comparable magnitude under certain conditions of diode operation. The first integral is the contribution to the power arising from the second-order correction to the voltage; it appears to have been neglected in previous discussions of diode theory. The second integral is the only contribution to the power which has been discussed in the literature. It may be interpreted as the diode "resistance" times the mean of the square of the alternating current.

The first integral is given by

$$\int_0^1 \left(\frac{d^2V}{d\mu^2}\right)_0 dt = 24\kappa \sum_{r=1}^{\infty} \frac{f_r f_{-r}}{\theta_r^4} \times\left[1-\frac{\theta_r^2}{4}+\frac{2}{\theta_r^2}(1-\cos\theta_r)-\frac{2\theta_r \sin\theta_r}{\theta_r^2}\right], \quad (8.8)$$

and the second by

$$\int_0^1 \left(\frac{dV}{d\mu}\right)_0 f(t)dt = 12\kappa \sum_{r=1}^{\infty} \frac{f_r f_{-r}}{\theta_r^4} \times[2(1-\cos\theta_r)-\theta_r \sin\theta_r], \quad (8.9)$$

where $\theta_r = \omega_r \kappa$.

Substituting Eqs. (8.8) and (8.9) into Eq. (8.7), one obtains

$$\left(\frac{d^2\langle P\rangle}{d\mu^2}\right)_0 = 24\kappa \sum_{r=1}^{\infty} \frac{f_r f_{-r}}{\theta_r^4}\left[1-\theta_r \sin\theta_r\left(1+\frac{2}{\theta_r^2}\right)+2(1-\cos\theta_r)\left(1+\frac{1}{\theta_r^2}\right)-\frac{\theta_r^2}{4}\right]. \quad (8.10)$$

The dimensionless quantity $\langle P\rangle$ is, therefore, to second-order

$$\langle P\rangle = \frac{3\kappa}{4}+12\mu^2\kappa \sum_{r=1}^{\infty} \frac{f_r f_{-r}}{\theta_r^4}\left[1-\theta_r \sin\theta_r\left(1+\frac{2}{\theta_r^2}\right)+2(1-\cos\theta_r)\left(1+\frac{1}{\theta_r^2}\right)-\frac{\theta_r^2}{4}\right], \quad (8.11)$$

or the average power per unit area is

$$\langle P'\rangle = \left(\frac{6a^4 I_0^5}{\epsilon_0^2}\frac{m}{e}\right)^{\frac{1}{2}}\left\{\tfrac{3}{4}+12\mu^2 \sum_{r=1}^{\infty} \frac{f_r f_{-r}}{\theta_r^4} \times\left[1-\theta_r \sin\theta_r\left(1+\frac{2}{\theta_r^2}\right)+2(1-\cos\theta_r)\left(1+\frac{1}{\theta_r^2}\right)-\frac{\theta_r^2}{4}\right]\right. \quad (8.12)$$

Llewellyn[3] has treated the case $f_r=0$, $r\neq 1$ and obtained the reduced form of Eq. (8.10), aside from constants, as the expression for the "resistance" of the diode. However, Eq. (6.11) shows that the "resistance" (the real part of the complex impedance at a single frequency) times the mean of the square of the single frequency alternating current does not give the power at this frequency. The additional terms are given by Eq. (8.8).

For a given frequency the ratio λ of the total second-order correction to the power to the "resistance" term is given by one of (8.12) to one of (8.10). Thus,

$$\lambda = \frac{1-\theta_r \sin\theta_r[1+(2/\theta_r^2)]+2(1-\cos\theta_r)[1+(1/\theta_r^2)]-\theta_r^2/4}{2(1-\cos\theta_r)-\theta_r \sin\theta_r}. \quad (8.13)$$

For small values of θ_r this becomes

$$\lambda = \tfrac{5}{6}+O(\theta_r^2) \quad (8.14)$$

and the large values of θ_r

$$\lambda = O(\theta_r), \quad (8.15)$$

where $O(x)$ refers to the usual order of greatness.[7] Thus the terms neglected in the Llewellyn treatment dominate the "resistance" term in the expression for the power for large θ_r. Even for small θ_r the correction is about 16 percent.

[7] E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, London, England, 1935), p. 11.

High Frequency Impedance of Low Pressure Gaseous Diodes

CHAI YEH
*Department of Electrical Engineering, University of Kansas, Lawrence, Kansas**

AND

E. L. CHAFFEE
Cruft Laboratory, Harvard University, Cambridge, Massachusetts

(Received March 20, 1950)

A simple theory of the lagging effect of the positive ions in neutralizing the space charge near the cathode of a gaseous diode is developed. By means of bridge measurement up to a frequency of 350 or 1000 kc per second, depending upon the type of tube used, the theory is checked experimentally in terms of the high frequency impedance of the diodes. Several constants of interest, such as the transit time and the lifetime of a positive ion, its effectiveness in neutralizing space charge, etc., can be deduced from the theoretical and experimental investigations.

I. INTRODUCTION

WHEN a conducting path of electricity possesses quantities which cause a lag and lead of an alternating current with respect to the voltage, the effect is ordinarily expressed in terms of inductive or capacitive reactance. This characteristic of a conductor may result from the inertia of the carriers of the charge wholly irrespective of any ordinary inductance or capacitance which the circuit may contain. Llewellyn[1] has treated the effects of the inertia of the electron at very high frequencies in modifying the conduction characteristics of high vacuum tubes. A similar inertia effect may be observed at low frequencies in the conduction through a gaseous vacuum tube. The reactance of the conduction path in this case results from the inertia of the positive ions. An interesting case of this type of conduction with inertia effects is described in this paper. A simple theory of the effect will first be presented and then experimental measurements on several soft tubes will be given to show that the theory at least approximately describes the effects observed. Certain fundamental quantities pertaining to the conduction process can be approximately determined from the experimental results.

The case to be treated is that of a cylindrical diode containing a trace of gas. The d.c. polarizing potential is adjusted to be slightly greater than the ionization potential of the gas and a small superimposed a.c. voltage is applied. The object is to find the equivalent a.c. resistance and reactance of the diode as a function of frequency.

The conditions assumed in the analysis are as follows: The static characteristic curve of the soft diode is of the type shown in Fig. 1, where $\bar{E}_i$ is the ionization potential. The increase in current above the normal space-charge limited current when the impressed potential exceeds $\bar{E}_i$ is assumed to be due to the partial neutralization of space charge near the cathode by positive ions. The current carried by positive and negative ions is assumed negligible.

The instantaneous anode current i at time t is not only a function of the instantaneous anode potential e but is also a function of the applied anode potential at a previous time $t-\tau$ denoted by $e_{t-\tau}$. The reason for the latter effect is that positive ions produced near the anode at time $t-\tau$ requires time τ to reach the space near the cathode where they neutralize some of the negative space charge. Expressed analytically,

$$i=f(e, e_{t-\tau}). \tag{1}$$

Equation (1) can be expressed in terms of n, the number of positive ions existing at time t in the cathode space. Then

$$i=F(e, n). \tag{2}$$

Since n does not depend upon e but is a function of $e_{t-\tau}$, e and n can be considered as independent variables. Then the differential of Eq. (2) is

$$di=(\partial i/\partial e)_n de+(\partial i/\partial n)_e dn=Kde+Gdn, \tag{3}$$

where $K\equiv(\partial i/\partial e)_n$, and is the slope of the $i-e$ curve for a constant number of positive ions, and $G\equiv(\partial i/\partial n)_e$, and is the effectiveness of partial neutralization of space charge by the positive ions at constant e.

II. CALCULATION OF THE NUMBER OF POSITIVE IONS

The evaluation of the current by Eq. (3) requires a knowledge of the number of positive ions, n at time t.

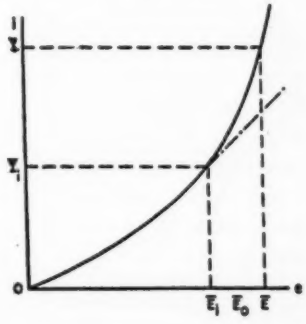


FIG. 1. Static characteristic curve of a gaseous diode. The dot-dash extension indicates the space-charge limited case in a high vacuum tube.

* Formerly, Cruft Laboratory, Harvard University, Cambridge, Massachusetts.

[1] F. B. Llewellyn, Proc. I. R. E. **21**, 1532 (1933), **22**, 947 (1934), **23**, 112 (1935).

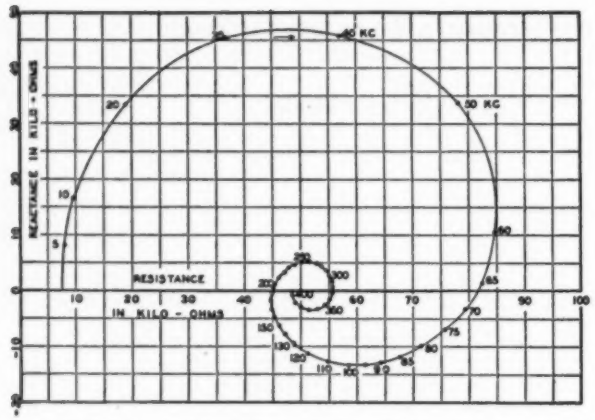


FIG. 2. Theoretical impedance curve of Type I, $B<K$, calculated with the following constants: $\bar{I}=0.2$ ma, $\bar{E}_0=1$ volt, $K=0.1955\cdot10^{-4}$ mhos, $P=0.1445\cdot10^{5}$ 1/volt²-sec., $\beta=0.401\cdot10^{5}$ 1/sec., $\tau=3.960\cdot10^{-6}$ sec. Frequency scale is spotted along the curve.

An expression for n will now be derived. Assume a cylindrical diode having an anode of radius a at potential e. The diode is assumed to contain a trace of gas which has an ionization potential $\bar{E}_i$. If $\bar{E}_i$ is slightly less than e, ionization occurs in a thin cylindrical space adjacent to the anode. The gas pressure is such that the mean free path of electrons is large compared with a.

Let $S(e_r)$ be the ionization coefficient at potential e_r, defined as the number of ions produced by one electron in traveling one centimeter at one-mm pressure. Then the number of ions produced by current i in time dt is

$$dN=(pi/10\epsilon)dt\int_{r_i}^{a}S(e_r)dr, \tag{4}$$

where p is the pressure in mm of mercury, i is the current in amperes, ϵ is the electronic charge in absolute e.m.u., and r_i is the radius corresponding to $\bar{E}_i$.

We shall assume a linear relation between e_r and r near the anode in the form $e_r=[(r-a)/g-1]e$, so that $dr=(g/e)de_r$. Furthermore, we shall assume a linear relation between $S(e_r)$ and e_r of the form $S(e_r)=b(e_r-\bar{E}_i)$. This latter relation is quite accurate for values of e_r not exceeding $2\bar{E}_i$. Using these relations, Eq. (4) becomes

$$dN=M(i/e)\cdot(e-\bar{E}_i)^2dt, \tag{5}$$

where

$$M=pgb/20\epsilon. \tag{6}$$

M is a constant for any given tube at constant pressure.

Equation (5) gives the number of positive ions produced in time dt near the anode. We shall indicate the mean transit time of a positive ion in traveling from anode to cathode by τ. Although some positive ions may strike the cathode, others miss the cathode and oscillate about it. We must therefore recognize a mean lifetime of positive ions in the cathode space. The decay of any elementary group dN in time t when near the cathode may be expressed by the relation

$$dn=dNe^{-\beta t}, \tag{7}$$

where β is the reciprocal of the mean lifetime of a positive ion as defined by Kingdon.[2] Multiplying Eq. (5) by (7) in which $t-\tau-\lambda$ is substituted for t, we have for the number of positive ions at time t,

$$n=\int_{-\infty}^{t-\tau}M(e_\lambda-\bar{E}_i)^2(i_\lambda/e_\lambda)e^{-\beta(t-\tau-\lambda)}d\lambda. \tag{8}$$

Now assume that $\bar{E}$ and $\bar{I}$ are the d.c. anode potential and anode current, where $\bar{E}$ is assumed to be just greater than $\bar{E}_i$. Let a small a.c. potential of angular frequency ω be superimposed on $\bar{E}$. Then

$$\begin{aligned}e&=\bar{E}+\Delta E\sin\omega t,\\ i&=\bar{I}+\Delta I\sin(\omega t-\phi).\end{aligned} \tag{9}$$

Substituting expressions (9) for e_λ and i_λ in (8) and

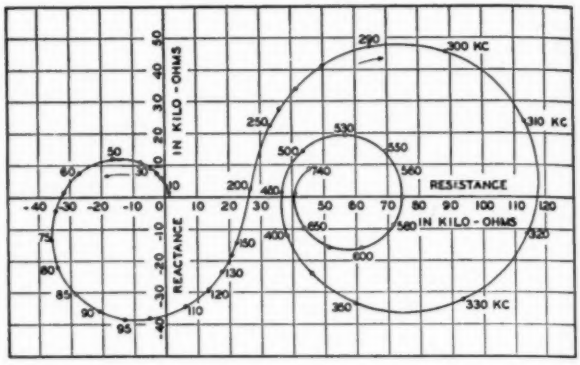


FIG. 3. Theoretical impedance curve of Type II, $B>K$, calculated with the same set of constants as the Type I curve except with $\bar{I}=1.5$ ma. The constants used in these calculations were actually determined from experimental data. The results check with each other very well up to 150 kc. Measurements at higher frequencies are lacking due to experimental difficulties.

integrating, we obtain n in the form $N+\Delta n$ plus high order terms. Retaining terms of fundamental frequency only and neglecting terms of the higher order of ΔE and ΔI, gives

$$\begin{aligned}\Delta n=\frac{\bar{E}_0M}{\bar{E}(\omega^2+\beta^2)^{\frac{1}{2}}}\{&\Delta E\bar{I}(2-\bar{E}_0/\bar{E})\sin[\omega(t-\tau)-\theta]\\&+\bar{E}_0\Delta I\cos[\omega(t-\tau)-\theta-\phi]\},\end{aligned} \tag{10}$$

where $\bar{E}_0=\bar{E}-\bar{E}_i$, $\theta=\tan^{-1}\omega/\beta$.

III. ALTERNATING CURRENT IMPEDANCE

We can now substitute (10) in (3) and equate to zero the coefficients of $\sin\omega t$ and $\cos\omega t$. Since $\sin\phi=X/Z$ and $\cos\phi=R/Z$, the values of R and X can be found and are

$$R=\frac{(K-BA)+(B-AK)\cos\delta}{K^2+B^2+2KB\cos\delta}, \tag{11}$$

$$X=\frac{(B+AK)\sin\delta}{K^2+B^2+2KB\cos\delta}, \tag{12}$$

[2] K. H. Kingdon, Phys. Rev. 21, 408 (1923).

$$Z^2=R^2+X^2=\frac{A^2-2A\cos\delta+1}{K^2+B^2+2KB\cos\delta}, \tag{13}$$

where

$$A=\frac{MG\bar{E}_0^2}{\bar{E}(\omega^2+\beta^2)^{\frac{1}{2}}}=\frac{P\bar{E}_0^2}{(\omega^2+\beta^2)^{\frac{1}{2}}}, \quad P=\frac{MG}{\bar{E}}, \tag{14}$$

$$B=\frac{MG\bar{E}_0\bar{I}}{\bar{E}(\omega^2+\beta^2)^{\frac{1}{2}}}\left(2-\frac{\bar{E}_0}{\bar{E}}\right)=\frac{P\bar{E}_0\bar{I}}{(\omega^2+\beta^2)^{\frac{1}{2}}}\left(2-\frac{\bar{E}_0}{\bar{E}}\right) \tag{15}$$

and

$$\delta=\omega\tau+\theta=\omega\tau+\tan^{-1}\omega/\beta. \tag{16}$$

It is interesting first to investigate the resistance and reactance at the two extreme frequencies, that is at zero and infinite frequencies. Substituting $\omega=0$, we have

$$\begin{aligned} X_0&=0, \\ \phi_0&=0, \\ R_0&=(1-A)/(K+B), \end{aligned} \tag{17}$$

where the zero subscript indicates the value at zero frequency. Here we find that the reactance is zero and the resistance is finite and positive as long as A is less than unity, which is true in practice. And for $\omega=\infty$, we have

$$\begin{aligned} X_\infty&=0, \\ \phi_\infty&=0, \\ R_\infty&=Z_\infty=1/K. \end{aligned} \tag{18}$$

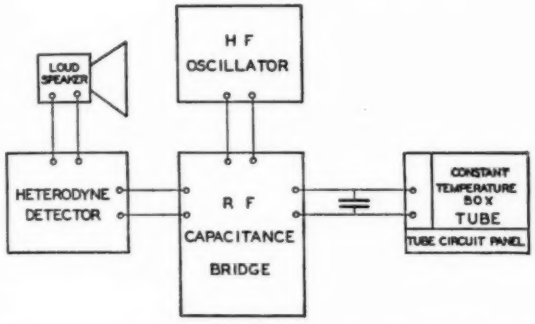


FIG. 4. Experimental set up for the measurement of high frequency impedance of a gaseous diode.

IV. THEORETICAL IMPEDANCE CURVES

Equations (11) and (12) represent the resistance and reactance components of the complex impedance of a gaseous diode. From Eqs. (14)–(16), it is seen that both the constants A and B and the transit angle δ are functions of frequency. An investigation of the frequency characteristics of this complex impedance can be made either by plotting R and X individually as a function of frequency, or by plotting the reactance against resistance. For the present discussion, the latter presentation is preferrable. As an aid in determining the shape of the impedance curve given by Eqs. (11) and (12), the intercepts on the R and X axis may first be found.

The intercepts on the X or vertical axis can be found by letting $R=0$, and solving Eqs. (11) and (12) simultaneously. The result is

$$X=\pm\left(\frac{1-A^2}{B^2-K^2}\right)^{\frac{1}{2}}, \ (R=0). \tag{19}$$

As A is always less than unity (which is required in order to have positive resistance at zero frequency), then $1-A^2$ is a positive quantity. If B^2 is greater than K^2, the term under the square root sign is real, and the positive and negative signs in front of it indicate the positive and negative intercepts on the X-axis corresponding to different frequencies. But if B^2 is less than K^2, the term under the square root sign becomes imaginary, which means that there are no intercepts.

The intercepts on the R or horizontal axis can be found by letting $X=0$ in Eqs. (11) and (12). Thus we have

$$R=(1+A)/(K-B), \quad (X=0). \tag{20}$$

It is seen that R can be either positive or negative according to whether K is greater or less than B, since both B and K are positive quantities. The above analysis leads to the discussion of two types of the impedance curves. They are very distinct in shape and are characterized by whether $B<K$ of $B>K$.

Type I, $B<K$.—In this case, the entire curve lies on the positive resistance side of the diagram, and there is no intercept on the X-axis. At low frequencies, the resistance is nearly constant while the inductive reactance increases with frequency. Above a certain frequency, both resistance and reactance change rapidly. The reactance changes from inductive to capacitive as $\sin\delta$ changes its sign. The resistance swings back and forth about the value of R at $\omega=\infty$ but remains positive. Both the resistance and reactance decrease in magnitude as ω increases and the curve spirals in a clockwise direction in towards the point corresponding to $1/K$ at $\omega=\infty$. This is indicated in Fig. 2, which is a plot of the theoretical curve of Type I for the following constants:

$\bar{I}=0.2$ ma $\quad \bar{E}_0=1$ volt, $\quad K=0.1955\cdot10^{-4}$ mhos,
$P=0.1445\cdot10^5$ 1/volt²-sec., $\quad \beta=0.401\cdot10^5$ 1/sec.,
$\tau=3.96\cdot10^{-6}$ sec.

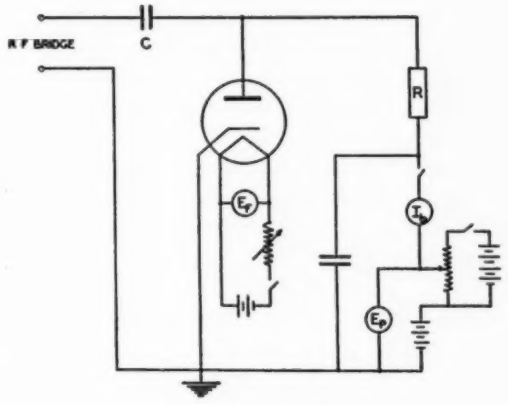


FIG. 5. Experimental tube circuit.

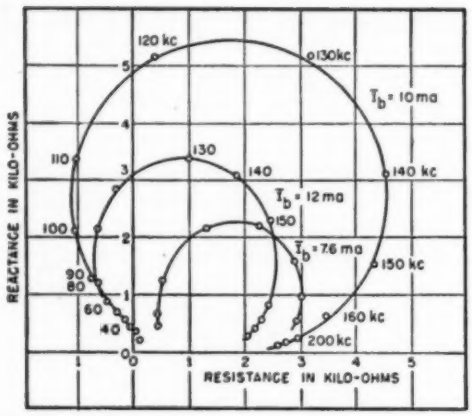


FIG. 6. Experimental impedance curve of a helium-filled diode.

Type II, for $B>K$.—For this type of curve, there are two intercepts on the X-axis which are $X=+[(1-A^2)/(B^2-K^2)]^{\frac{1}{2}}$ and $-[(1-A^2)/(B^2-K^2)]^{\frac{1}{2}}$ corresponding

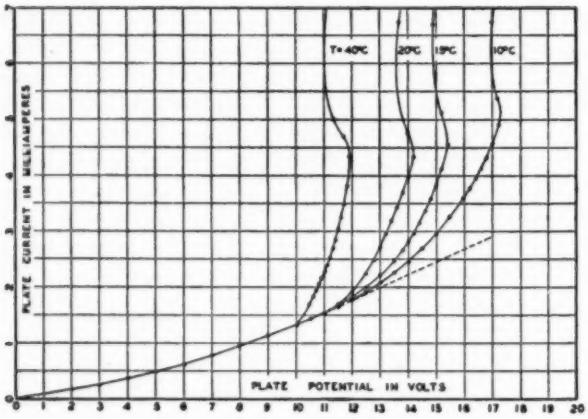


FIG. 7. Static characteristic curve of the experimental tube RK-100, containing mercury vapor.

to different frequencies. There is a negative intercept on the R-axis, which corresponds to a negative resistance. As K is a constant and a positive quantity, while B is a positive number but diminishes in magnitude as frequency increases, it is obvious that as the frequency increase further, B decreases and finally becomes less than K. Then the curve takes the form of Type I.

At very low frequencies, the resistance is positive and practically constant. As the frequency becomes greater, it decreases and becomes negative at certain frequencies and then returns to positive values progressing in a counter-clockwise direction along a spiral. For higher frequencies, an inflection point is reached and the curve spirals in a clockwise direction and the resistance varies in the manner of Type I. The entire curve of Type II is plotted in Fig. 3 for the same set of constants used for Type I curve except that $\bar{I}=1.5$ ma instead of 0.2 ma.

V. EXPERIMENT

In the theory developed so far, we made use of four unknown constants which may be determined experi-

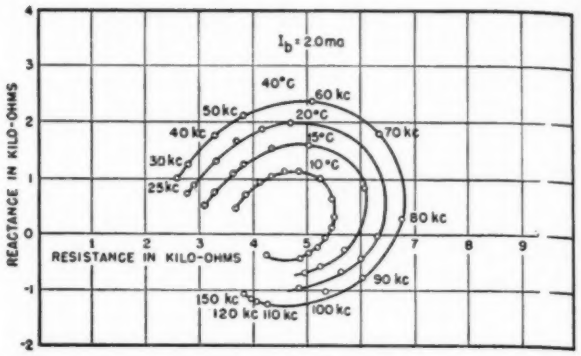


FIG. 8. Effect of gas pressure on the impedance curve at low anode current. Tube contains mercury vapor.

mentally from the intercepts of the experimental curves. These are τ, β, K, and $P(=MG/\bar{E})$. It is better to use P instead of A and B, since both A and B are derivable from P.

The general arrangement of the apparatus used for measuring the tube impedance is shown in Fig. 4. The entire set up consists of a high frequency bridge, a high frequency oscillator, a sensitive detector, and the tube circuit. The tube, whose impedance characteristic is to be measured, is mounted in a constant-temperature oil bath controlled by a thermostat. The complete tube circuit is shown in Fig. 5. In this circuit, C is a blocking condenser so as to isolate the d.c. anode potential from the bridge circuit. R is one of the several non-inductive resistances having ranges from several hundred to several hundred thousand ohms. The anode potential can be varied by battery taps and more precisely by the potential divider. The whole system is enclosed in a well-grounded aluminum box.

Before taking any measurement on the high frequency bridge, a d.c. static characteristic curve of the tube is taken. The compensation method is adopted for the measurement on the high frequency bridge. The anode potential of the tube to be measured is adjusted to a proper point on the d.c. characteristic. The output of the oscillator is kept low in order to conform to the assumption that the a.c. component is very much smaller than the d.c. values. For each frequency, the bridge is balanced by adjusting both the air condenser and the parallel resistance on the bridge. Two sets of readings are necessary; one with the anode potential of the tube on, and the other with it off. The first set of readings gives the tube admittance in parallel with that of the circuit, while the second set of readings gives the circuit admittance alone. The observed change in resistance and capacitance is that due to the tube impedance. The resistance and reactance thus measured are plotted and the intercepts on the real and imaginary axes are determined. Equations (17), (19), and (20) may then be used to evaluate the constants P, β, and K and Eq. (16) is used to calculate τ. A curve of Type II is required to provide enough intercepts to determine the four constants.

VI. RESULTS AND DISCUSSIONS

Measurements on several gaseous tubes were made. Some of the tubes were filled with helium gas, others with mercury vapor. The results obtained from the helium-filled tubes indicate that the time of flight of the positive ions is extremely small. A frequency range of from 30 to 1000 kc, is used in this experiment. The impedance curve thus obtained is plotted in Fig. 6. It is noticed that the reactance of this tube remains inductive even at extremely high frequencies. The resistance is negative during a small portion of the frequency range. But as the anode current is varied, both the resistance and reactance remain positive. The impedance curve of this tube is essentially of Type I, but has characteristics of the transition type between those of I and II where there are two positive values of X for zero values of R and where the spiraling is always clockwise. We were not able to increase the anode current $\bar{I}$ or the gas pressure sufficiently to obtain the anticlockwise spiral of Type II. The anode current saturates at a slightly higher operating point. It is therefore impossible to determine the constants of this tube.

The next group of tubes were filled with mercury vapor. Several special diodes were constructed by the Raytheon Manufacturing Company. They are cylindrical diodes with indirect-heated cathodes. The cathode radius is 0.075 cm and the anode radius is 1 cm. The d.c. characteristic curve of one of these tubes is shown in Fig. 7. The curves are taken at different temperatures of the bulb which determines the vapor pressure inside the tube. Above an anode current of 4.5 ma, a greenish glow starts and an abrupt increase of anode current results which is accompanied by a decrease in potential as is shown in the plots. All measurements were made at an anode current well below the glow discharge.

Figures 8 and 9 are taken for the purpose of illustrating the effect of varying the gas pressure on the impedance characteristics of the tube. In Fig. 8, the anode current is held constant at 2 ma, while in Fig. 9, it is held at 4 ma. It is noticed that for low anode current, the changing of gas pressure merely changes the dimension of the impedance curve without materially affecting its position or shape. These are shown as concentric curves in Fig. 8, and are essentially of Type I. For higher anode current, the variation of the gas pressure shows the evolution of the impedance curve from Type I to Type II. Figures 10 and 11 are taken for the purpose of illustrating the effect of changes of anode current on the impedance curve of the tube. Two sets of data were taken, one for low pressure and the other for high pressure. At low pressure, the shape of the impedance curve does not change very much and the increase of the anode current merely moves the curves towards lower resistance and reactance as shown in Fig. 10. For high pressures, the evolution of the impedance curves from Type I to Type II as the anode current increases is apparent as shown in Fig. 11. This time, the factor that affects B is the anode current. Hence both the anode current and the gas pressure are responsible for the change of shape of the impedance curve. For an anode current greater than 4 ma, and the bulb temperature higher than 40°C, the hiss inside the tube becomes so

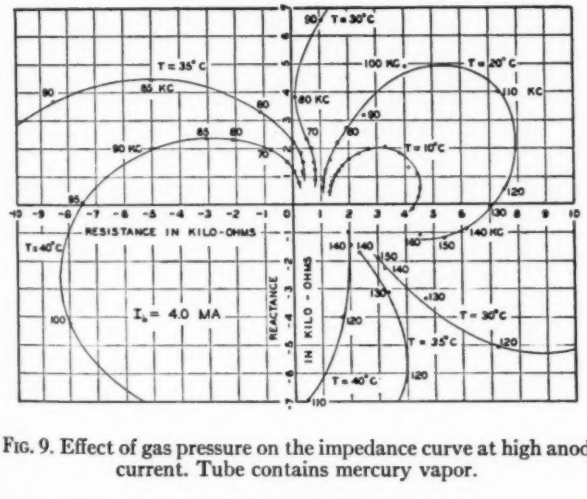


FIG. 9. Effect of gas pressure on the impedance curve at high anode current. Tube contains mercury vapor.

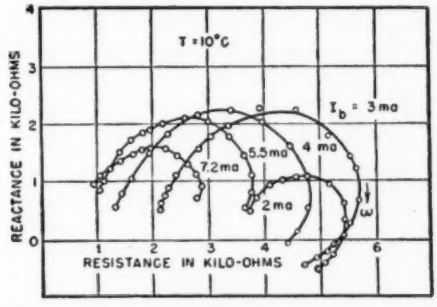


FIG. 10. Effect of anode current on the impedance curve at low gas pressure. Tube contains mercury vapor.

intense that bridge balance is difficult and inaccurate. Figure 12 is another plot of one of the curves in Fig. 11 (for $\bar{I}=4.0$ ma) together with the theoretical curve calculated from the constants preciously determined. The theoretical and experimental data are carried far into the higher frequency region in order to show the change over from an anticlockwise to clockwise rotation of the spiral as in the Type II curve.

For a typical operation corresponding to an anode current of 4.0 ma, and a mercury-vapor pressure corresponding to 40°C, the following constants were determined:

Typical operation condition	Intercepts from experimental data	Calculated constants
$\bar{E}=11.86$ volts	$R_0=350$ ohms	$K=0.239\cdot10^{-4}$ mhos
$\bar{E}_0=1.48$ volts	X (at $R=0$) $=1200$ ohms	$\beta=0.164\cdot10^{6}$ 1/sec.
$\bar{I}=4.0$ ma	R (at $X=0$) $=-7500$	$P=0.268\cdot10^{8}$ 1/V²-sec.
$T=40$°C	$\omega_1=2\pi50\cdot10^{3}$	$\tau=3.28\cdot10^{-6}$ sec.
	$\omega_2=2\pi90\cdot10^{3}$	

As an example of the application of these constants, the following computations are made. Knowing $1/\beta$, the

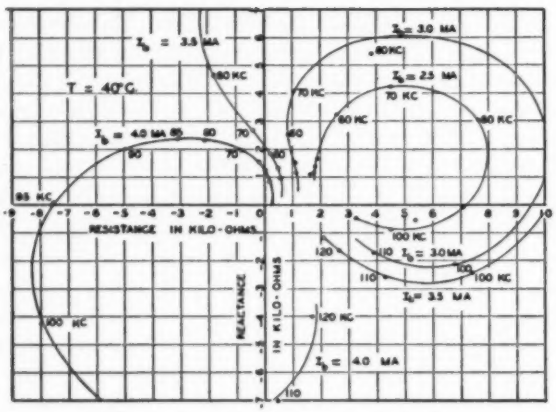


FIG. 11. Effect of anode current on the impedance curve at high gas pressure. Tube contains mercury vapor.

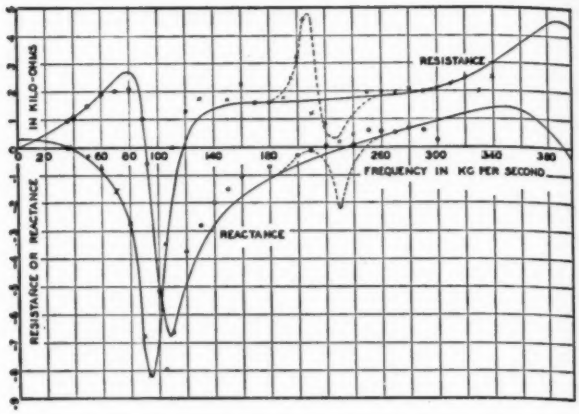


FIG. 12. Theoretical and experimental check of a typical operation: $\bar{E}=11.86$ volts, $\bar{E}_0=1.48$ volts, $\bar{I}=4.0$ ma, $T=40°C$. Theoretical curves are drawn in solid lines, the circles and crosses are the experimental values. Circles for reactance, crosses for resistance. The experiments and calculations were carried out to higher frequencies, up to 340 kc on this tube. The dotted curve indicates the region where hiss is excessive.

mean lifetime of the positive ions, we may proceed to calculate the average number of times that the positive ions circulate around the cathode provided that we know the equivalent radius of the circular path of the circulating ions. From the consideration that the centrifugal force of the moving ions are balanced by the force acting on the ions due to the electric field applied between the electrodes, and making use of γ, which is a function of r as given in Langmuir's paper,[3] the radius of the circulating path is

$$r_a=2\gamma^2/(d\gamma^2/dr_a). \tag{21}$$

This equation can best be solved by cut and try method. The solution is found for a cathode radius of 0.075 cm and is $r_a=0.080$ cm. The number of times that an ion circulates around the cathode is given by $N_a=v/2\pi r_a\beta$, where v is the velocity of the positive ions and is given by $v=(2E\epsilon/m_+)^{\frac{1}{2}}$. For a mercury ion, the average number of times that the ions circulate around the cathode for the conditions stated is 4.08 or roughly 4 times.

The constants P and β enable us to calculate approximately the average number of electrons that are neutralized by a single positive ion under the stated condition. From Eq. (7), it is seen that the rate of disappearance of positive ions is $dn/dt=-\beta N$. In order to maintain the steady state condition, the rate of supply of positive ions should be equal to βN, which is maintained by the increment in the positive ion current. If the increment in positive ions is proportional to the increment in the positive ion current, it is found that

$$\frac{\text{Number of electrons}}{\text{Number of positive ions}}=\frac{G}{\epsilon\beta}=\frac{P\bar{E}}{M\epsilon\beta}. \tag{22}$$

For an operating temperature of 313°K, which corresponds to a mercury vapor pressure of 0.004 mm of Hg, and the constant of proportionality $b=1.5$, $g=0.925$, the average number of electrons that are neutralized by a single positive ion under this operating condition is equal to 7480, or roughly 8000.

[3] I. Langmuir, Phys. Rev. **21**, 435 (1923), **22**, 347 (1923).

The Propagation of Plastic Deformation in Solids*

THEODORE VON KARMAN AND POL DUWEZ
Pasadena, California
(Received March 27, 1950)

The stress wave caused by a longitudinal impact at the end of a cylindrical bar has been analyzed in the case where the impact velocity is large enough to produce plastic strain. The theory gives a method for computing the stress distribution along the bar at any instant during impact. It is shown that for a given material, there is a critical impact velocity such that when subjected to a tension impact with a velocity higher than the critical, the material should break near the impacted end with negligible plastic strain.

An experimental investigation was made concurrently with the theoretical study. Some of the most significant experimental results are presented in this paper.

I. INTRODUCTION

THE testing of impact strength of materials previous to the discovery of the laws of propagation of plastic deformation used essentially two different methods: bending of notched bars, and tension or compression of short cylindrical specimens. The first method presumes a certain standardization of the shape of the specimens and tries to establish certain relative merit numbers for the impact resistance of various materials by measuring the energy absorbed by the standardized specimens. The second method is somewhat more ambitious; its aim is the determination of the energy absorption by unit volume of the material, in the case of impact, and the comparison of this amount of energy with the work absorbed in static experiments. These tests, especially the tests of the second type, have shown that for moderate impact velocities the work absorbed in impact is, in general, somewhat larger than the work absorbed in static experiments; there were some indications that this tendency is reversed as the impact velocity increases. However, the observed drop in the amount of absorbed work was not accompanied by a decrease in elongation and no explanation was offered for this phenomenon.

As the European war started, the attention of military and other engineers concerned with the construction of bomb resisting structures was focused on the impact problem. The senior author became interested in the problem of resistance of structures to impact loads beyond the elastic limit in a conference held at the National Academy of Sciences early in 1940, under the chairmanship of Dr. V. Bush. He conceived the idea of the plastic wave in order to obtain a method for theoretical calculation of the energy absorbed by various structures, as columns, beams, plates, etc. The theory of elastic wave propagation was known. It was evident from this theory that, for any especially brittle material (i.e., materials which break at their elastic limit), a critical impact velocity exists for which fracture must occur at the point of application of the impact, so that the rest of the structure has no chance to participate in the energy absorption. It appeared that if a theory of the propagation could be developed, such a theory would reveal the existence of a critical velocity for structures loaded beyond their yield strength.

The basic theory and preliminary experiments supporting the theory, at least in its fundamental conclusions, were carried out during the year 1941. The senior author presented the fundamentals of the theory in a paper sent for publication in the Proceedings of the National Academy of Sciences in December, 1941. It was found, however, that the paper might have some bearing on problems relating to National Defense. The publication was therefore delayed, but the study was described in a classified NDRC report, A-29, "On the Propagation of Plastic Deformation in Solids," February 1942. The experimental work carried out previously by the junior author was published at the same time in the classified NDRC report, A-33, "Preliminary Experiments on Propagation of Plastic Deformation." Subsequently more complete theoretical investigations on the subject were carried out by the senior author, H. F. Bohnenblust, D. H. Hyers, and J. Charyk. The main content of these investigations will probably be published in the near future.

II. LONGITUDINAL IMPACT AT THE END OF A BAR EXTENDING TO INFINITY

The propagation of elastic strains in a cylindrical bar subjected to tension impact had already been analyzed in 1807 by Thomas Young, who established the proportionality between the elastic strain and the velocity of impact. The value of the elastic strain ϵ is in this case given by the formula

$$v_1 = c_0 \epsilon,$$

in which v_1 is the velocity of impact and c_0 is the velocity of propagation of an elastic deformation. Young had pointed out in his paper that when the velocity of impact is above a certain critical value a plastic strain is initiated near the point of impact. Since Young's work, it seems that the interest of the problem has not been fully recognized, and no systematic attempt has been made to compute the stress and strain caused by an impact beyond the elastic limit of the material. In the present paper such a treatment is provided for the case of longitudinal impact.

* Presented at the Sixth International Congress for Applied Mechanics, Paris, France, September 22–29, 1946.

Consider a rod or wire extending from $x=-\infty$ to $x=0$ and assume that the endpoint at $x=0$ is suddenly put in motion with a constant velocity v_1. Let the stress-strain relation for the material be given by a function of the form $\sigma=\sigma(\epsilon)$ where σ is the stress and ϵ the strain. Stresses that depend on the time-rate of strain will be neglected; for, with the exception of the case of extremely high velocities, such stresses are small in comparison with the stresses that depend on the strain itself. To be sure, the relation $\sigma=\sigma(\epsilon)$ holds only for the first deformation of the material beyond the elastic limit; in the case of load reversal, another functional relation, which takes the hysteresis into account, has to be used. The lateral contraction of the material—that is, the contribution of the lateral contraction to the kinetic energy—is neglected in the following calculation.

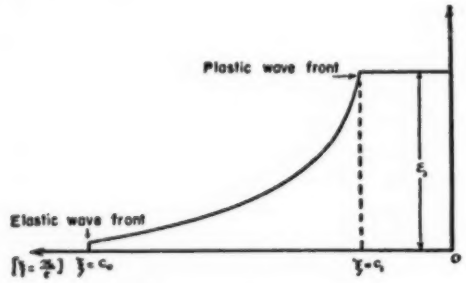


FIG. 1. Schematic representation of ϵ *versus* ξ.

With these simplifications, the equation of motion for an element of the rod or wire can be written in the form

$$\rho\frac{\partial^2 u}{\partial t^2}=\frac{d\sigma}{d\epsilon}\frac{\partial \epsilon}{\partial x}, \tag{1}$$

where u is the displacement of the element in the longitudinal direction, ρ is the density of the material, and t is the time.

It should be pointed out that in Eq. (1) the process of deformation is considered from the Lagrangean point of view. An arbitrary cross section of the bar is determined by a coordinate x, which gives its distance from the origin in the unloaded state. The stress is defined as the ratio of the internal reaction between two portions of the bar to the initial cross-sectional area of the bar. The strain is numerically equal to the change of length of a portion of the bar which, in the initial unloaded state, was of unit length. If u is the displacement of a cross section, the strain ϵ is equal to $\partial u/\partial x$. It is not assumed that ϵ is small in comparison with unity.

Since $\epsilon=\partial u/\partial x$, Eq. (1) can also be written in the form

$$\rho(\partial^2 u/\partial t^2)=T(\partial^2 u/\partial x^2), \tag{2}$$

where $T=d\sigma/d\epsilon$ is the modulus of deformation, elastic or plastic. The quantity T is considered to be a given function of the strain. The boundary conditions are $u=v_1 t$ for $x=0$, and $u=0$ for $x=-\infty$.

It is easily seen that a solution of the form

$$u=v_1[t+(x/c_1)] \tag{3}$$

with an arbitrary value of the velocity of propagation c_1, satisfies Eq. (2) and the boundary condition at $x=0$. For this solution the strain ϵ is constant and is equal to v_1/c_1.

A second solution is obtained by putting

$$T/\rho=x^2/t^2. \tag{4}$$

Since $T[=d\sigma/d\epsilon]$ is a given function of ϵ, Eq. (4) represents a solution for which ϵ is a function only of the variable $\xi=x/t$.

Assume $\epsilon=f(\xi)$; then the displacement u has the form

$$u=\int_{-\infty}^{x}\frac{\partial u}{\partial x}dx=\int_{-\infty}^{x}f(\xi)dx=t\int_{-\infty}^{\xi}f(\xi)d\xi \tag{5}$$

since $dx=td\xi$. By differentiation one readily obtains

$$\begin{aligned}\partial^2 u/\partial t^2&=(\xi^2/t)f'(\xi)\\ \partial^2 u/\partial x^2&=(1/t)f'(\xi)\end{aligned} \tag{6}$$

and substitution of Eqs. (6) in Eq. (2) shows that one of the two equations,

$$\rho\xi^2=T \tag{7}$$

or

$$f'(\xi)=0 \tag{8}$$

must hold. Equation (8) leads to the solution expressed by Eq. (3), whereas Eq. (7) gives the solution of Eq. (4).

The complete solution is obtained as follows:

(a) For $|x|<c_1 t$ the strain ϵ is constant and equal to ϵ_1;

(b) For $c_1 t<|x|<c_0 t$, where c_0 is the velocity of propagation of the elastic wave,

$$T(\epsilon)=\rho(x^2/t^2) \tag{9}$$

(c) For $|x|>c_0 t$, $\epsilon=0$.

The distribution of ϵ as a function of ξ is shown schematically in Fig. 1. The value of T for small values of ϵ—that is, within the elastic limit—is equal to E, Young's modulus of elasticity for the material. The elastic wave propagates with the velocity $c_0=(E/\rho)^{\frac{1}{2}}$. Between the plastic wave front, which is propagated with the velocity c_1, and the elastic wave front, the strain is variable, since every strain-increase from ϵ to $\epsilon+d\epsilon$ proceeds with a velocity equal to the specific value of $c=(T/\rho)^{\frac{1}{2}}$, corresponding to the strain ϵ.

The main problem is to determine the velocity c_1 of the plastic wave and the maximum strain ϵ_1 as a function of the velocity of impact v_1.

From Eq. (5) it follows:

$$v_1=\frac{u(0,t)}{t}=\int_{-\infty}^{0} f(\xi)d\xi \tag{10}$$

and by changing variables

$$v_1=-\int_0^{\epsilon_1}\xi d\epsilon$$

(See Fig. 4).

Thus, upon substituting for ξ from Eq. (7), Eq. (10) takes the form

$$v_1=\int_0^{\epsilon_1}(T/\rho)^{\frac{1}{2}}d\epsilon. \tag{11}$$

Since T is a given function of ϵ, Eq. (11) determines ϵ_1 as a function of v_1.

If the deformation remains within the elastic limit, $T=E=$constant, and $v_1=\epsilon_1 c_0$, then $v_1=\epsilon_1(E/\rho)^{\frac{1}{2}}$. Hence, the stress σ is given by

$$\sigma_1=E\epsilon_1=\rho v_1 c_0. \tag{12}$$

Equation (12) is universally used for the calculation of the stress produced in an elastic body subjected to an impact velocity v_1. It appears that Eq. (11) replaces Eq. (12) in the case of a deformation beyond the elastic limit. If the stress remains within the elastic limit, there are two regions: for $|x|<c_0 t$, $\sigma=\rho v_1 c_0$ and for $|x|>c_0 t$, $\sigma=0$. In the case of plastic deformation, there are two fronts. Beyond the front of the elastic wave, $\sigma=0$; between the fronts of the elastic and plastic waves, σ increases gradually from $\sigma=0$ to a maximum value, $\sigma=\sigma_1$; and behind the front of the plastic wave, σ has the constant value σ_1, corresponding to a total strain ϵ_1 —elastic plus permanent—where ϵ_1 is given by Eq. (11).

For most materials $d\sigma/d\epsilon$ approaches zero for large values of ϵ, and at some particular value of ϵ, the material breaks. Hence the integral constituting the right-hand member of Eq. (11) has a maximum value, and one obtains a critical value of the velocity v_1. It can be expected that an impact with a velocity larger than v_1 will cause an instantaneous breakdown of the material.

III. EXTENSION OF THE THEORY; BAR OF FINITE LENGTH

The characteristic parameters that define the state of strain and motion of an element are the strain $\epsilon[=\partial u/\partial x]$, the stress σ, and the velocity of the element $v=\partial u/\partial t$. The equation of motion for the element is

$$\rho(\partial v/\partial t)=\partial\sigma/\partial x$$

or, using the definition of the velocity of propagation,

$$\partial v/\partial t=c^2(\partial\epsilon/\partial x). \tag{13}$$

On the other hand, from the relations $v=\partial u/\partial t$ and $\epsilon=\partial u/\partial x$ it follows that

$$\partial v/\partial x=\partial\epsilon/\partial t. \tag{14}$$

Equations (13) and (14) are equivalent to the differential Eq. (2).

The coordinate x of the element and the time at which the magnitude of the strain of this element is equal to ϵ and its velocity is equal to v can be considered as functions of ϵ and v. Then Eqs. (13) and (14) take the form

$$\begin{aligned}\partial x/\partial\epsilon&=c^2(\partial t/\partial v)\\ \partial t/\partial\epsilon&=\partial x/\partial v.\end{aligned} \tag{15}$$

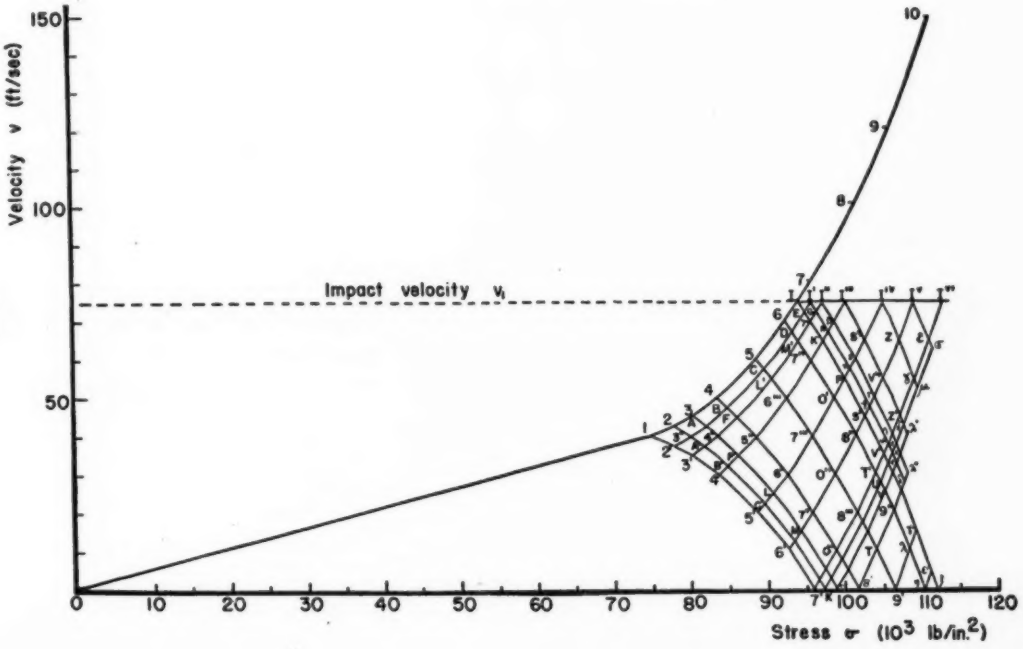


FIG. 2. Velocity *vs.* stress diagram for heat treated carbon steel.

Equations (15) become more symmetrical by the introduction of the function $\phi = \int_0^{\epsilon} c(\epsilon) d\epsilon$ and we obtain

$$\partial x/\partial \phi = c(\partial t/\partial v)$$
$$\partial x/\partial v = c(\partial t/\partial \phi). \tag{16}$$

The process of propagation can be represented in two planes. The first, with x and t as coordinates, is the Lagrangean or physical plane; the second plane is the velocity plane, in which ϕ and v are used as coordinates. The system of Eqs. (16) has fixed characteristics in the $v\phi$ plane. Hence, a relatively simple graphical method can be developed for the solution of a broad class of impact problems in the plastic range; for example, the case of a bar of finite length with one fixed end and hit at the free end. In Figs. 2 and 3 are typical (σ, v) and (x, t) diagrams for a specimen of heat treated carbon steel subjected to an impact velocity of 75 ft./sec.

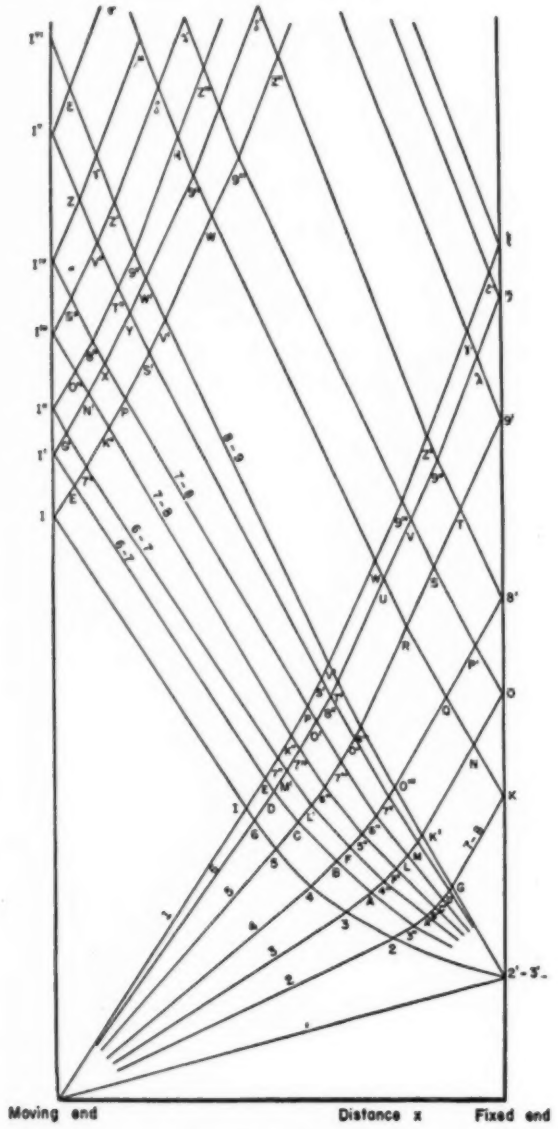


FIG. 3. Lagrange diagram for specimen of heat treated carbon steel. Impact velocity 75 ft./sec.

IV. EXPERIMENTS ON THE PROPAGATION OF PLASTIC STRAINS IN A LONG SPECIMEN SUBJECTED TO TENSION IMPACT

1. Experimental Technique

The first experiments on the propagation of plastic strain in tension were made in view of verifying the three following points resulting from the theory: (1) the existence of a plastic wave front of given amplitude; (2) the relation between the amplitude ϵ_1 of the plastic wave front and the velocity of impact v_1 (Eq. (11)); (3) the distribution of the plastic strain between the plastic and the elastic front as given by Eq. (9). The specimens used in these experiments were annealed copper wire about 100 inches long and 0.071 inch in diameter. The specimens were mounted vertically in a special testing machine built by Dr. D. S. Clark at the California Institute of Technology. The impact was produced by a hammer guided between two rails and accelerated by prestretched rubber bands. The maximum velocity attainable was approximately 200 ft./sec. The velocity was measured by a suitable electric device and a cathode-ray oscillograph. On every specimen equidistant marks were made with one inch spacings. After the test, the plastic strain was determined by measuring the displacement of each mark. Experiments were also performed in compression impact. In this case, the specimens were cylinders 12 inches long and about $\frac{3}{8}$ inch in diameter.

To control the duration of impact, the following device was employed: the bottom end of the wire is attached to a rigid piece A, as shown in Fig. 4. A vertical rod B, resting on the bottom frame of the machine, fits loosely into the tubular part of A. When a hammer H hits the piece A, the specimen elongates until A reaches the rod B. The piece A contains a circular notch N, and the rim of A breaks off at this notch after A has traveled the distance D and comes to rest on the rod B. The purpose of this arrangement is to allow the hammer to continue to move downward and also to dissipate some of its remaining energy. However, no kinetic energy is transferred to the specimen after A reaches B. The time of impact is therefore the distance D divided by the velocity of the hammer during the process of elongation.

The static stress-strain curve of one of the copper wires tested in impact was used to compute the values of $d\sigma/d\epsilon$ as a function of ϵ and the values of the velocity of propagation c, as given by the formula $c = [1/c(d\sigma/d\epsilon)]^{\frac{1}{2}}$, were then computed and a curve of c *versus* ϵ was obtained (Figs. 5 and 6). From this curve, the value of the quantity v_1 (the velocity of impact corresponding to a plastic front of amplitude ϵ_1) was calculated for each value of ϵ_1 (Fig. 7). The largest value of v_1 is obtained for the largest possible value of (16 percent in this case) and is equal to 150 ft./sec. The impact velocity of 150 ft./sec. is therefore the "critical velocity" for this material. An impact with a higher velocity

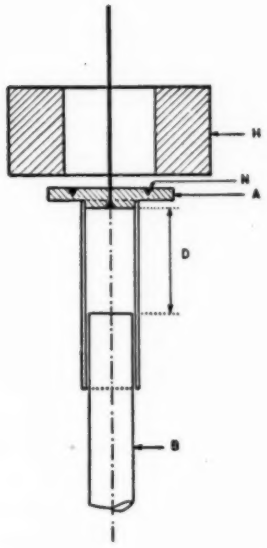


FIG. 4. Experimental device used to stop the impact after a given deformation of the specimen has been reached.

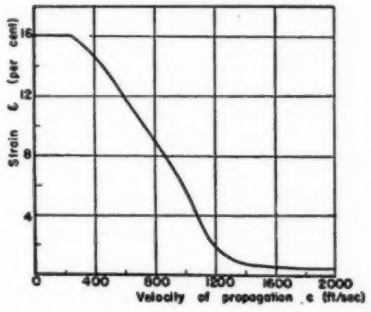


FIG. 5. Strain ϵ *versus* velocity of propagation c as calculated from the static stress-strain curve.

must produce an instantaneous breakdown of the specimen.

As mentioned above, marks were made on the specimen at intervals of one inch. The distance between the origin and each mark was measured before and after the test. If x_n denotes the distance from the origin to a mark n before the test, and if x_n' is the distance between the two points after the test, the difference $x_n'-x_n$ represents the total elongation of the specimen between these two points. The values of $x_n'-x_n$ were then plotted *versus* x. The slope at any point of the curve so obtained gave the value of the permanent strain at that point. This way of measuring the strain has been found to be the most practical one. It was not necessary to make the distinction between permanent strain and actual strain, because the elastic recovery was relatively too small to be taken into consideration.

2. Results of the High Velocity Impact Tests

The results are given in three separate subsections—(i), (ii), and (iii)—in order to show to what extent they agree with the three principal theoretical results which were expressed in Section II.

(*i*) *Existence of a Plastic Wave Front*

The first series of tests was made to establish that the amplitude of the plastic front is a function of the velocity of impact alone. This amplitude remains constant while the elastic front and the plastic front travel along the specimen. In these particular experiments, the velocity of impact was always 92.50 ft./sec., but the durations of the impact varied. To control the duration of impact the two pieces A and B, shown in Fig. 4, were placed at a proper distance apart. The curves in Fig. 8 give the distribution of the strain along the specimen. They indicate clearly that a plastic front of a given amplitude ϵ_1 is revealed by the experiments, as predicted from the theory.

(*ii*) *Relation Between the Velocity of Impact and the Amplitude of the Plastic Wave Front*

In this series of tests, the velocity of impact was varied from one test to the other. The total elongation was not necessarily the same for all the tests. The stopping device was adjusted in such a manner that during the impact the plastic front traveled a distance of between 20 and 40 inches. Figure 9 gives the distribution of the strain along the specimen in each case. The experimental values of ϵ_1 corresponding to each velocity tested are listed in Table I. For comparison with the theory, the experimental values are plotted as points in Fig. 7, whereas the solid curve represents the result of the theoretical computation. The agreement between the experimental results and the predetermined curve is fairly good. The highest velocity of impact used was 171 ft./sec. For that velocity, the specimen broke within the first inch, indicating that a velocity of 171 ft./sec. is above the critical velocity, as predicted by the theory. It must be pointed out that a considerable reduction of area was observed, which indicates that the rupture was not brittle. In addition, a plastic strain of relatively small intensity (less than

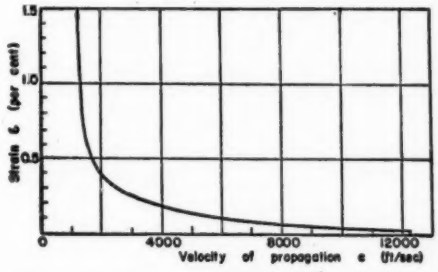


FIG. 6. Continuation of the curve of Fig. 5 up to the elastic point.

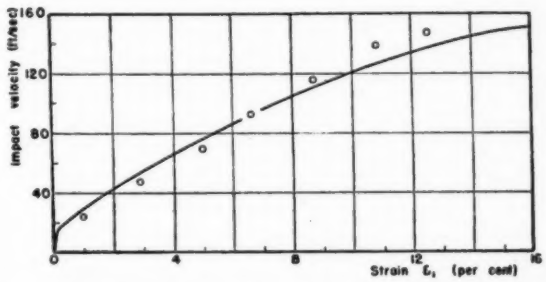


FIG. 7. Variation of the strain ϵ_1 with the impact velocity v_1.

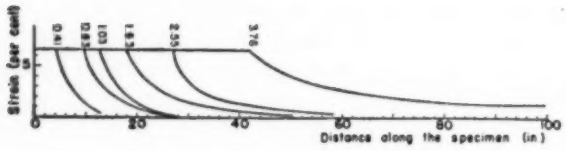


FIG. 8. Distribution of the strain ϵ along the specimen. Velocity of impact 92.5 ft./sec. The duration of impact is indicated on each curve in msec.

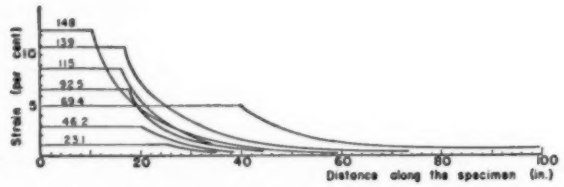


FIG. 9. Strain distribution curves for annealed copper specimens subjected to different impact velocities, indicated on each curve in ft./sec.

two or three percent) was measured along the wire over a distance of about 20 inches from the point of rupture. It is probable that a plastic wave started to propagate along the specimen during the time required to produce rupture.

(*iii*) *Shape of the Plastic Wave and Velocity of Propagation of the Plastic Front*

The distribution of the plastic strain along the wire between the plastic and elastic fronts may be calculated from Eq. (9), using the curves of Figs. 5 and 6. The dashed curves in Fig. 10 represent theoretically computed amplitude distributions of a plastic wave after a time interval equal to 0.83 msec.; the assumed velocity of impact, v_1, is 92.5 ft./sec. and ϵ_1 is 6.6 percent. The solid curve in Fig. 10 represents plots of the measured distribution of the strain. It is seen that the experimental and theoretical curves deviate considerably. This deviation is due to some inaccuracy in the value of the duration of impact, and especially to the perturbating effects brought by the sudden stopping of the impact. When the hammer which pulls the end of the wire with a constant velocity is stopped, the portion of the specimen that is in motion is not stopped instantaneously. During the deceleration period, the kinetic energy stored in the specimen is transformed into strain energy, and the total extension is larger than the distance D through which the hammer has been in contact with the specimen.

TABLE I. Results of impact tests on 80 inches long annealed copper specimens.

Impact velocity (ft./sec.)	Uniform strain (percent)	Duration of impact (msec.)	Velocity of propag. of uniform strain (ft./sec.)
23.1	1.0	1.70	1230
46.2	2.9	1.47	1130
69.4	5.0	3.37	1000
92.5	6.6	1.63	910
115.0	8.7	1.55	860
139.0	10.8	1.94	730
148.0	12.5	1.40	630
171.0		Rupture	

The measured shape of the plastic wave given in Fig. 10 is greatly influenced by this "stopping effect" and cannot logically be compared with the theoretical curve that represents the shape of the wave at a certain instant during impact. The stopping effect was analyzed theoretically by the authors and especially by Drs. F. Bohnenblust and J. Charyk, who collaborated in the work reported in this paper. These investigations will be published separately. They materially reduced the variance between computed and measured values.

The reflection of plastic waves at the fixed end of a specimen was also investigated theoretically and experimentally. A typical result of this investigation is shown in Fig. 11, in which a measured strain distribution curve obtained in a compression impact test is compared with a computed curve. The two curves are not in complete agreement, but the character of the observed strain distribution curve is accounted for by the theory. It may also be pointed out that the deviation between the two curves certainly reflects some influence of the strain rate, showing that the assumption of a stress-strain curve independent of the rate of strain is not entirely justified.

A very noticeable effect of the rate of strain on the elastic limit has been observed for iron specimens subjected to both tension and compression impact. The most striking result of impact tests on iron specimens

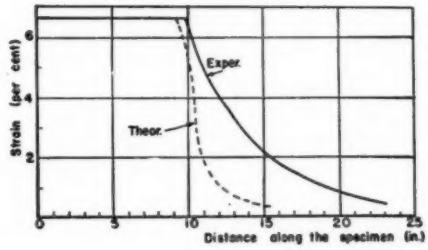


FIG. 10. Theoretical and experimental strain distribution curves for annealed copper specimen. Impact velocity 92.5 ft./sec. Duration of impact 0.83 msec.

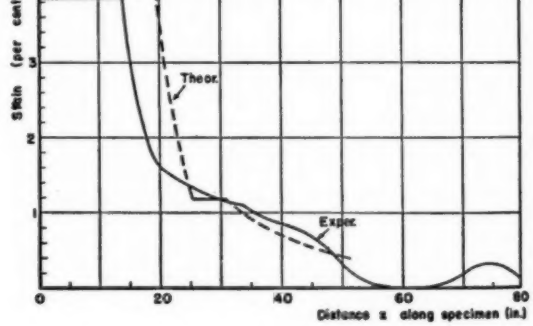


FIG. 11. Experimental and theoretical strain distribution curves for a copper wire 0.08 inch in diameter. Impact velocity 62.5 ft./sec., duration of impact 1.3 msec.

was that no plastic strain occurred near the moving end until the stress wave reached a value of the order of three times the static yield stress. The measured strain distribution curves also showed a different character, as shown in Figs. 12 and 13. It can be observed that in the case of compression the permanent set is concentrated in the neighborhood of the ends of the specimen and the center is free from plastic strain. In the tension tests there were intervals of permanent sets alternating with undisturbed regions. This is probably due to irregularities in the specimens whose influence was

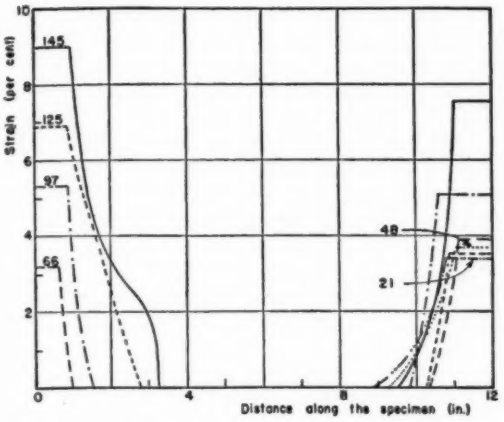


FIG. 12. Strain distribution curves for annealed mild steel specimens subjected to compression with different impact velocities, indicated on each curve in ft./sec.

enhanced by the unstable character of the tension process.

3. Force-Time Relation at the Fixed End of a Specimen Subjected to Tension Impact

An important contribution of the theory of plastic wave propagation presented in this paper is a clear explanation of the dependence on time of the force measured at the fixed end of a specimen subjected to tension impact. Force *versus* time diagrams recorded at the fixed end of the specimen have been interpreted by previous investigators in the field of high velocity impact testing as dynamic stress-strain curves. The transformation of the time axis into a strain axis infers that the strain is uniform all along the specimen at any instant during impact. The theory of strain propagation shows that for impact velocities greater than approximately 10 ft./sec., this hypothesis is not justified, and a stress-strain relation cannot be obtained in such a simple way. In order to show this, the force acting at the fixed end of a specimen subjected to tension impact was computed as a function of time by the theory outlined in Section III. This force was also measured in a series of experiments made with different types of specimens. The force was recorded by means of a resistance sensitive strain gauge in conjunction with a cathode ray oscillograph. Two records obtained for copper specimens tested with an impact velocity of 100 ft./sec. are reproduced in Fig. 14, together with a theoretical stress *versus* time diagram. The agreement between theoretical and experimental curves is very satisfactory as far as their shape is concerned. The experimental curve, however, seems to be shifted to the left. This discrepancy could be explained by assuming that any strain has a velocity of propagation slightly higher than that based on the static stress-strain curve. This gives further evidence of the influence of the rate of strain on the relation between stress and strain. In any case, the interpretation of such curves as those of Fig. 14 as dynamic stress-strain curves is completely fictitious.

4. Plastic Strain Propagation and High Velocity Impact Tests

High velocity tensile impact tests have been reported by several investigators.[1–6] Some of the results[1] have shown that the energy absorbed before rupture decreased when a certain impact velocity was reached.

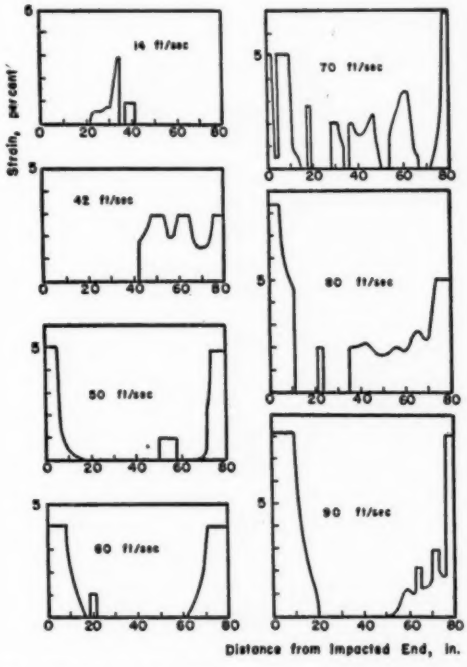


FIG. 13. Strain distribution for annealed iron wire in tension for different impact velocities.

However, the observed drop in the amount of absorbed word was not accompanied by a decrease in elongation

[1] H. C. Mann, Proc. A.S.T.M. **36**, 85 (1936).
[2] D. S. Clark and G. Datwyler, Proc. A.S.T.M. **38**, 98 (1938).
[3] D. S. Clark, *The Influence of Impact Velocity on the Tensile Characteristics of Some Aircraft Metals and Alloys*, NACA Technical Note No. 868.
[4] M. Manjoine and A. Nadai, Proc. A.S.T.M. **40**, 822 (1940). Trans. A.S.M.E., J. App. Mech. **8**, A-77 (June, 1941).
[5] DeForest, MacGregor, and Anderson, Metals Technol. **8**, 1 (December, 1941).
[6] E. R. Parker and C. Ferguson, Trans. A.S.M. **XXX**, No. 1, 68 (March, 1942).

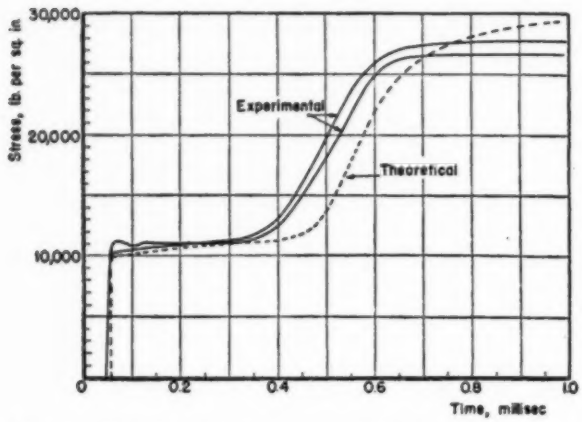


FIG. 14. Experimental and theoretical stress-time curves in tension for fixed end of annealed copper. Gage length eight inches; impact velocity 100 ft./sec.

and no explanation was offered for this phenomenon. Other investigators did not observe any drop in energy absorption, even with impact velocities as high as 200 ft./sec.

After the discovery of the laws of propagation of plastic deformation, it became evident to the authors that both the elongation and the energy should decrease more or less suddenly when the critical velocity is reached. The first tests made with long copper wires and described in Section III clearly established the existence of a critical velocity, as predicted by the theory. The rupture of the specimen, however, at velocities above the critical, was still of the ductile type, and some appreciable strain due to necking could still be measured near the end of the specimen subjected to impact. The elongation due to necking and the energy absorbed by the necking process are relatively important when the test specimen is short, or rather when the ratio between the length and the diameter of the specimen is small. Systematic experiments indicated that, in order to observe a definite drop in both energy and total elongation beyond the critical velocity, specimens should have a ratio of length to diameter at least equal to 20. These results furnished a clear explanation why a critical velocity was not observed by previous investigators, who made high velocity impact tests on short specimens.

One of the fundamental questions in high velocity impact tests is the influence of the rate of strain. The investigations of the authors clearly showed that such an influence exists. In fact, whereas the general character of the impact process was fairly well described by a theory neglecting the influence of the rate of strain, there was always a certain systematic discrepancy pointing to such an influence. However, it should be emphasized that in high speed impact tests the rate of strain varies from one point to the other and is a function of time. Computations resulting from the theory of wave propagation have shown that the average rate of strain calculated on the basis of uniform strain is far from the actual rate of strain, even when the velocity of impact is as low as 10 ft./sec. It is therefore not logical to utilize tension impact tests to study the influence of the rate of strain on the properties of metals.

Finally, it appears that for materials having a yield point like soft iron, the interpretation of high velocity impact tests is made rather difficult because of the phenomena reported in Section IV-2. The authors believe that the understanding of these phenomena requires a deeper insight into the whole question of plastic deformation and the physics of solids.

V. PRACTICAL APPLICATIONS

As previously stated, the theory of propagation of plastic deformation was developed for the understanding of the resistance of structures against impact. Hence the first application which appears to be important is the development of rules which enable the designer to compute the amount of energy which can be absorbed by a given structure. It has been shown quite clearly for the first time that the amount of energy is a function of the impact velocity, with a rather sudden decrease of the energy absorption beyond a critical velocity. The theory has already been extended to beams subjected to transverse impact. A class of ballistic problems which have not been sufficiently discussed from the point of view of plastic strain should be mentioned. Certain aspects of the problem of penetration of projectiles through metal plates are connected with plastic wave propagation. It seems, for example, that the notion of critical velocity could lead to the discrimination of the cases in which the penetrating projectile merely pushes aside the material from those in which a plug is separated from the back of the plate. Finally, the notion of plastic strain propagation might bring about a better understanding of some anomalies encountered in high speed machining.

A Kerr Cell Camera and Flash Illumination Unit for Ballistic Photography

H. F. Quinn, W. B. McKay, and O. J. Bourque
Canadian Armament Research and Development Establishment, Valcartier, Quebec, Canada
(Received March 30, 1950)

This paper describes an experimental model of a combination Kerr cell camera and flash illumination unit designed specifically for the photography of projectiles under conditions of strong ambient illumination. Specific cases wherein such conditions are found are considered and the applicability of the Kerr cell as a secondary protecting shutter is described. A description is given of a pulse forming network—pulse transformer circuit designed to apply an operating voltage pulse to a Kerr cell of thirty-six (36) kilovolts amplitude and one point nine (1.9) microseconds duration. Limitations inherent in the present equipment are considered together with proposals for overcoming these defects in future models. Typical results are presented, including (a) a picture taken with the unit showing a twenty-millimeter shot at the time of emergence from the muzzle of a gun, and (b) a picture of a thirty-seven millimeter shot taken shortly after penetration of a metal target plate.

THE "single shot" photography of projectiles in flight by means of a high intensity flash lamp of the Egerton type is now a well-established technique in most ballistic research laboratories. In this technique, the exposure time of each picture is determined by the effective photographic duration of the particular flash lamp employed and is normally of the order of a microsecond. Such an effective exposure is entirely adequate to "stop" all but the highest speed projectiles in flight. It is a conventional procedure to use a good camera of the Eastman "Speed Graphic" type with a shutter setting of 1/100 second, the shutter opening being suitably controlled by electrical circuits of various degrees of complexity to ensure that the shutter is open for some short interval before, while, and for a short interval after the projectile is in the field of view of the camera. Auxiliary electrical circuits are provided whereby the projectile, in passing some predetermined position in the field of view, triggers the flash lamp and the required picture is taken. As it is a common procedure to employ very fast film in these applications, some means must be provided to protect the film from fogging caused by ambient daylight even with the short exposure of 1/100 second and, in many cases, both the camera and the light source are positioned in a darkened hut through which the projectile passes via suitably placed apertures in the walls. A typical photograph of a conventional shell taken under such conditions is shown in Fig. 1.

Frequently, however, it is required to obtain pictures of projectiles under conditions where the subject is either completely or partially surrounded by a glowing medium of considerable luminosity. Examples of such cases are:

(a) A projectile emerging from the muzzle of a gun and surrounded by high intensity luminous gases (muzzle flash).

(b) A projectile striking an armor plate and surrounded by an atmosphere of glowing, incandescent particles of metal which have been stripped from the plate by the projectile impact.

A particularly good example of the ambient flash illumination produced in (b) is illustrated in Fig. 2, a photograph taken at this laboratory of an aluminum-capped projectile attacking an armor plate. It should be observed that the extraneous illumination present in such cases would not only tend to screen the projectile from the external photographic flash but also, by its own inherently high luminosity and relatively long duration (of the order of 1/100 second) would very completely fog the film during the 1/100 second opening of the mechanical shutter. Such total fogging will be observed in Fig. 2 in the regions of maximum interest to the terminal ballistician namely at the point of impact (A) on the front surface of the plate and at the point of emergence (B) at the rear. An attempt to employ a faster mechanical shutter would be entirely ineffective: the fastest mechanical shutter is one of the focal plane type, models of which have been built with quoted shutter speeds of 1/5000 second. Not only would such a shutter prove almost impossible to synchronize with the external flash source but would also fail to protect the film from fogging due to the ambient illumination from regions A and B even with the above short exposure. It is, then, necessary to resort to a relatively inertia-less, non-mechanical shutter. Some wartime investigations

FIG. 1. Conventional "Microflash" photograph of a shell in flight.

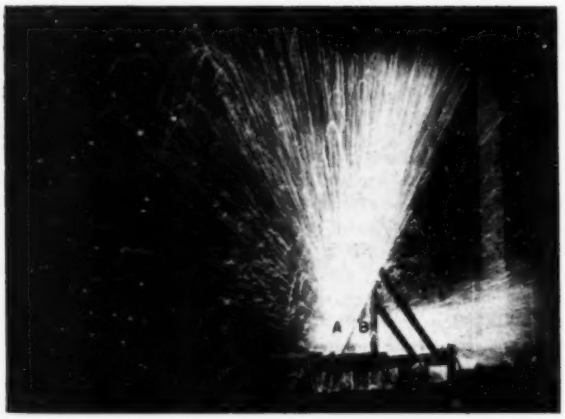


FIG. 2. Ambient flash illumination produced by an aluminum-capped projectile striking an armor plate.

performed in Germany and also in Great Britain suggested the possibility of utilizing the practically inertialess Kerr Cell electro-optical shutter in series with a conventional mechanical shutter to exclude the effect of the ambient flash. Security limitations do not, unfortunately, permit reference either to these investigations or to the results obtained therein. It would, of course, be imperative to effect a precise synchronization between the opening of the Kerr shutter and the initiation of the external flash lamp to insure that the illumination caused by the latter was fully utilized. At the end of the external flash, the Kerr shutter would close to exclude the ambient flash and, finally, the mechanical shutter would close to eliminate daylight fogging of the film noting particularly that the crossed polaroid sheets commonly used in most present day Kerr cells are reasonably transparent in the red portion of the spectrum. We desire to report on a preliminary, experimental model of a simple, self-contained unit incorporating both a Kerr cell camera and its associated flash illumination source for use in such applications as have been described above. The use of an artificial transmission line type of pulse-forming network together with a pulse transformer to obtain a substantially rectangular voltage pulse of two microsecond duration and 36 kilovolt peak amplitude for operating the Kerr cell is believed to be a new feature. It is further desired to point out some major disadvantages in the present equipment and to indicate how these may be overcome in a future model.

THEORY OF THE KERR CELL SHUTTER AND PREVIOUS APPLICATIONS

The theory and method of operation of the Kerr cell electro-optical shutter has been extensively discussed in the physics literature of the past fifty years; probably the most modern and satisfactory treatment being that due to Zarem.[1] It is not, therefore, proposed to include this information in the present report. It is, however, of interest to mention some of the uses to which the Kerr shutter has been applied by previous investigators and to consider the various aspects in which the present usage of the cell differs from them. Accordingly, Table I has been constructed in which several of the former applications have been listed together with the duration of the effective photographic exposure obtained in each case.

It is to be noted that, in all but the present application, the subjects to be photographed themselves provided all the necessary illumination to enable reasonably good pictures to be obtained notwithstanding the high light absorption losses in the active fluid (usually nitrobenzene) of the cell. The present usage of the Kerr shutter differs markedly from the others in that the useful light must be provided from an external, high intensity flash source, the light flash being accurately

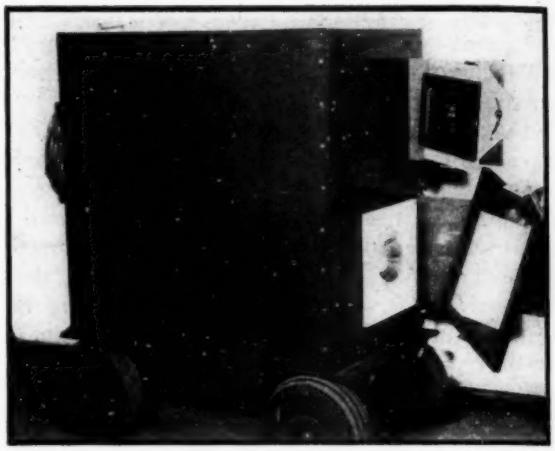


FIG. 3. Kerr cell camera and flash illumination unit.

synchronized with the opening of the shutter. Further, in the present application, the effective opening of the shutter need not be set precisely equal to the desired exposure time, provided it is at least sufficient to admit the external flash illumination and it is the effective photographic duration of the light flash that determines the exposure and not the opening of the Kerr shutter. Finally, the present application necessitates the use of a Kerr cell of much greater aperture and angular field of view than those employed by the above investigators; this, in turn, requires the use of much higher driving voltages in order to obtain a satisfactory opening of the shutter.

DESCRIPTION OF THE APPARATUS

A. Optical Arrangements

A photograph of the complete unit is shown in Fig. 3. The mirrors mounted on the front of the unit enable the latter to be suitably protected from flying fragments in the case of armor plate investigations and from the effect of muzzle blast in applications where it is desired to photograph projectiles emerging from a gun muzzle.

[1] Zarem, Marshall, and Poole, Elect. Eng. 68, No. 4, 282 (1949).

The general arrangement of equipment as used in the latter application is shown in Fig. 4 wherein the unit is seen in position beside the gun: the lower mirror enables the substantially parallel beam of light from the flash lamp to be trained on the muzzle while the upper mirror allows the camera to view the projectile on emergence. Initiation of both the Kerr camera and the flash lamp is provided by electrical contact between the emerging projectile and a stout wire tightly strung some one-half inch clear of the muzzle.

The camera proper consists of a "Baco" mounting and bellows together with a suitable adapter which permits the mounting of a $f/2.5$ "Aero Ektar" lens having a focal length of 8 inches. In this preliminary, experimental model the mechanical shutter was dispensed with and all tests were carried out at night inside a darkened gun shed. Suitable synchronization of a conventional, mechanical shutter from the gun itself would, however, present no great problem such an arrangement having been frequently employed by this laboratory in the past.

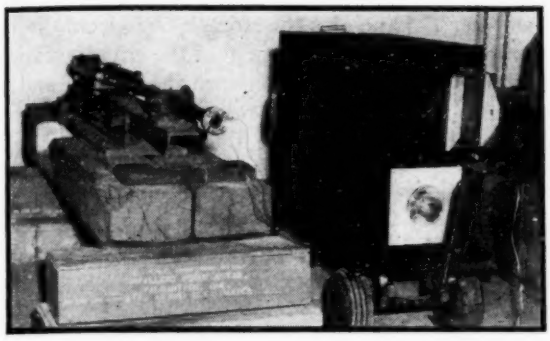


FIG. 4. Experimental arrangement for photography of a projectile at point of emergence from a gun muzzle.

Insofar as the dimensions of the electrodes are concerned, the Kerr cell follows exactly a German wartime model specified to operate at a peak voltage pulse of 36 kilovolts. As it was anticipated that considerable investigation might have to be carried out before the optimum design of the cell was finalized, the use of the more convenient (and much more expensive) all-glass cell was avoided and, to date, all the experimental cells have been constructed from one-inch Bakelite sheet and ordinary quarter-inch plate glass. All joints have been carefully sealed with ceresin wax and care has been taken to remove the nitrobenzene from the cell after each trial to eliminate the small but noticeable solution of this wax in the fluid. Type J Polaroid film pressed between the cell windows and external glass plates has been used to date for the linear polarizers but it is intended shortly to replace this by Type H film laminated in glass. The particular Kerr cell used in the first successful trial of the unit is shown in Fig. 5. The cell is placed immediately in front of the camera lens in the unit. It is not possible at the present time to calculate

TABLE I.

Author	Application	Effective exposure microseconds
Beams[a]	Study of the successive appearance of spectral lines in spark discharges.	0.001 to 0.1
Dunnington[b]	Photographic study of the early stages of the spark discharge.	0.002
Anderson[c]	Measurement of the velocity of light.	—*
Froome[d]	Fundamental studies of cathode spot formation in the arc discharge.	0.1
Zarem[e]	Fundamental study of the process of exploding wires.	0.04
Present authors	Ballistic photography.	2.00

* In this application, the Kerr cell was utilized as a light modulator valve.
[a] J. W. Beams, J. Opt. Soc. Am. 5, 597 (1926).
[b] F. G. Dunnington, Phys. Rev. 38, 1506 (1931).
[c] W. C. Anderson, Rev. Sci. Inst. 8, 239 (1937).
[d] D. K. Froome, J. Sci. Inst. 25, No. 11, 371 (1948).
[e] See reference 1.

accurately the theoretical maximum transmission of the Kerr cell when used with the present light source, the spectral distribution of energy in the flash being unknown. It is, however, planned to photograph the spectrum of the discharge lamp using Super XX film (the film actually used in applications of the equipment) and to substitute in the transmission formula the appropriate value of the Kerr constant corresponding to the spectral region of maximum light intensity. Some idea of the performance of this particular cell can, however, be obtained by substitution in the transmission equation of values for both the Kerr constant and the

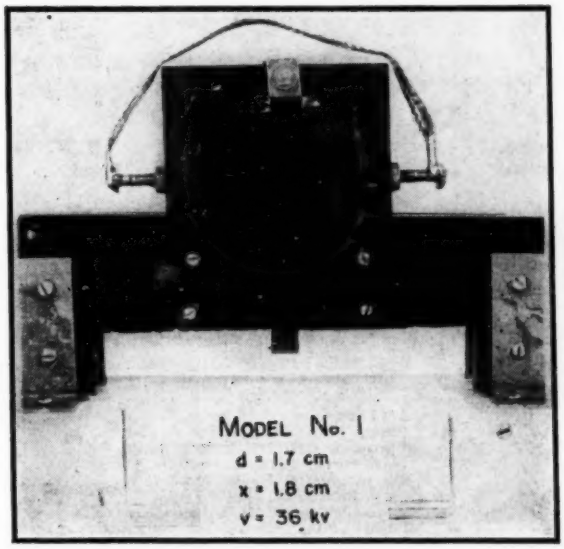


FIG. 5. First experimental Kerr cell.

absorption coefficient for a wave-length reasonably well in the center of the optical spectrum: let the following numerical values be used:

Kerr constant for nitrobenzene at λ
$= 5460$ angstroms $= 410 \times 10^{-7}$.

Absorption coefficient at $\lambda = 5460$ angstroms $= 1/15$.

The transmission equation, corrected for absorption in the cell, can be given in the form:

$$\text{Percent transmission, } T = 50 \exp -\rho \times \sin^2\left(\frac{\pi \times BV^2}{d^2}\right), \quad (1)$$

where x (cm) is the axial length of the cell plates, B the Kerr constant, d (cm) the plate separation, ρ the absorption coefficient and V (statvolts) the applied voltage. Substituting the above numerical values and the cell dimensions we obtain the result, $T = 37.1$ percent, the losses inherent in the linear polarizers being neglected.

In the present apparatus, the components available permitted operation of the cell only up to the originally specified limit of 36 kilovolts. Provided the dimensions x and d be fixed, it is apparent from Eq. (1) that a maximum possible percent transmission of 44.3 percent would be obtained for $V = 41.9$ kilovolts.

B. Electrical Circuits

The complete electrical circuit for the Kerr camera and its associated flash lamp is given in Fig. 6. Referring to Fig. 6, it will be seen that, when the high voltage supply is energized, one side of the main spark gap S charges to $+9$ and the other to -9 kilovolts. These potentials also appear across the flash energy storage capacitors C-4, C-5 and the pulse-forming network, PFN. On shorting the grid of the mercury thyratron V-3 to ground, a negative pulse of 4.5 kilovolts amplitude appears at the center electrode this being sufficient to trigger the main gap. Proper initiation of the main spark was found to be ensured by adding the capacitor C-3 across the gap. As soon as the gap fires, a total potential of 18 kilovolts appears across the flash lamp, a General Electric type FT-125, and this breaks down immediately, simultaneously discharging the pulse-forming network through the primary of the pulse-transformer T-5 and also the storage capacitors through the lamp itself. This arrangement of the circuit ensures a consistent synchronization between the camera and the flash lamp. A suitable purely resistive load of 640 ohms is connected in parallel with the Kerr cell across the

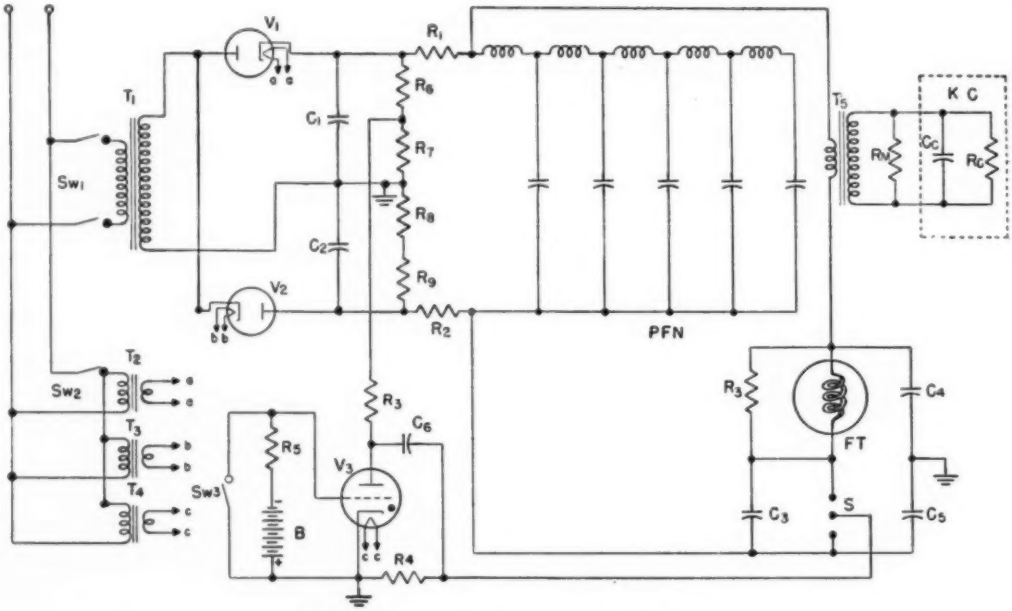


FIG. 6. General circuit diagram of the unit.

Referring letters	Components	Referring letters	Components
B	4–45 volts dry cell in series.	R_4	20 M at 20 kv.
C_1, C_2	Storage Capacitors, 0.25 μf at 25 kv.	R_5	10 K.
C_3	Capacitor, 0.0011 μf at 20 kv, mica.	R_6, R_7	10 M at 10 kv.
C_4, C_5	Flash energy storage capacitors, 1 μf at 10 kv.	R_8, R_9	10 M at 10 kv.
C_6	Capacitor, 0.0011 μf at 20 kv, mica.	Rm	Matching resistor—640 ohms.
FT	Flash tube, G.E. Type FT-125.	S	Control spark gap.
$K.C.$	Equivalent electrical circuit of Kerr cell.	Sw_3	Initiating contact switch.
PFN	Pulse-forming network—Zn = 40 ohms, $T = 1.9$ microsec.	T_1	Transformer—115, 7000 v.a.c.; 20 ma.
Cc	Capacitance of Kerr cell, $\simeq 55$ $\mu\mu$f.	T_2, T_3, T_4	Transformer, filament.
Rc	Resistance of Kerr cell, $\simeq 60$ K.	T_5	Transformer, pulse, step-up 1:4.
R_1, R_2	10 M at 10 kv.	V_1, V_2	High vacuum rectifier, type: 8013.
R_3	20 M at 20 kv.	V_3	Mercury thyratron, British CV-22.

secondary of the pulse transformer to match the transformer primary to the characteristic impedance of the pulse-forming network (40 ohms, in this particular case). While it will be noted that the voltage pulse at the transformer secondary cannot begin to rise until the initiation of the flash lamp, and that, by this means an approximate synchronism of the cell opening and the light flash is achieved, it is immediately apparent that the "rise time" to peak of the secondary voltage of a pulse transformer is quite considerable particularly when the latter works into an essentially capacitive load such as the Kerr cell (capacity of the order of 55 micro-microfarads). Lack of a suitable high speed single transient type oscilloscope has prevented the determination of the "rise time" in the present transformer. The photographic evidence is, however, that the "rise time" is, unfortunately, long enough to limit the useful light to that from the tail-end of the flash. There is no apparent simple means of eliminating this objectionable feature in the present equipment but, in future models, it is proposed to provide separate triggering of the cell and the lamp together with a precision time delay in between the two trigger pulses whereby the secondary voltage of the pulse-transformer will be given adequate time to rise to peak before the initiation of the flash. The fact that the photographic results obtained to date with this relatively crude apparatus have been so promising would seem to augur well for the success of the new equipment which is presently under construction.

It is important to note that the pulse transformer application is particularly appropriate to the present problem wherein a cell opening of one or two microseconds duration is the shortest required. The production of a pulse transformer with a four or five to one step-up ratio capable of providing a shorter voltage pulse of, for example, 0.1 microsecond to a capacitive load of some 60 micro-microfarads in parallel with the normal stray winding capacity (itself normally of the order of 40 mmf) would be a practical impossibility. Since, however, recent pulse transformers (with step-up ratios as great as one to seventeen) have been designed to handle two or three microsecond pulses, the use of a suitable pulse transformer should enable the operation of Kerr cells of much larger aperture with a consequent improvement in the photographic properties of the system as a whole. Furthermore, by the use of larger aperture Kerr cells, it should be possible to effect a substantial reduction in the capacity of the cells with a consequent reduction in the required power rating of the pulse transformer.

Concerning the several components used in the circuit, the pulse-forming network is of conventional design having been salvaged from a wartime radar set: it is rated at 25 kilovolts, has a specified pulse-width of 1.9 microseconds and a characteristic impedance of 40 ohms. The energy storage capacitors *C*-4 and *C*-5 were chosen from the standpoint of availability and insulation rating rather than for their particular suitability for flash duty; in fact they are high voltage filter capacitors. These capacitors will, of course, be replaced in future models.

Considerable difficulty was encountered throughout our investigations in the search for a suitable flash tube which would absorb the necessary power without disintegrating and which would also trigger from a suddenly applied surge voltage. While the FT-125 lamps

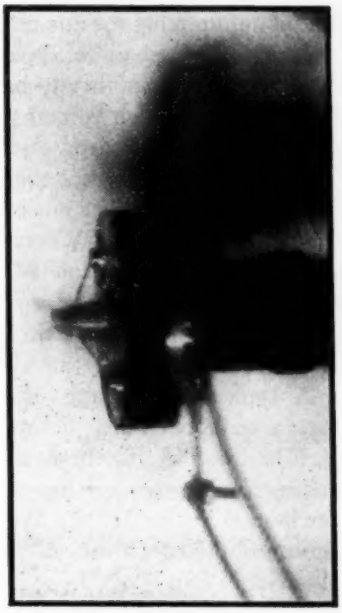


FIG. 7. Kerr cell photograph of a 20-mm shell emerging from gun muzzle.

finally selected are required to operate at a power level greatly in excess of their normal commercial rating, these lamps have given excellent results to date and, after some hundred flashes, still show no apparent signs of deterioration.

The pulse transformer was constructed in this laboratory: as it was initially desired merely to examine the applicability of the electrical circuit, no great effort was made in the construction of the device to minimize either the leakage inductance or the winding capacitance. A silicon steel core obtained from a surplus radar pulse transformer was utilized. The primary winding consists simply of two 40 turn windings each wound on one leg of the core and connected in parallel, the secondary consisting of two 80 turn windings each wound over one primary section and connected in series to give a voltage transformation ratio of 1 to 4. The transformer is immersed in oil in an old automotive battery case. Since the first promising results were obtained with the present equipment, orders have been placed with a manufacturer for a suitable pulse-forming

network having a characteristic impedance of 200 ohms and a pulse-width of 2 microseconds together with an associated pulse transformer having a 1 to 5 step-up ratio. This combination should enable a cell to be pulsed with a peak voltage up to the limit of 50 kilovolts.

EXPERIMENTAL PROCEDURE AND INITIAL RESULTS

The first significant and promising results were obtained in an application of the equipment wherein it was desired to obtain pictures of a twenty (20)-millimeter cannon shot just emerging from the muzzle of the gun for the purpose of determining the amount of propellant gas leakage past the driving band of the projectile. The experimental arrangement has already been illustrated in Fig. 4. The filament and high voltage power supplies were switched on and the thyratron grid connected to the contact wire at the muzzle; since the gun itself was grounded, the emerging projectile connected the grid to ground on contact with the wire. When these preliminary arrangements had been completed, the gunshed was darkened and the protecting slide removed from the film holder. Immediately after firing the gun, the slide was replaced in order to protect the film from stray illumination. A typical photograph of the emerging projectile (one of a series of twenty pictures obtained) is shown in Fig. (7) while Fig. (8) shows the opaque gas cloud surrounding the projectile when the latter had just cleared the muzzle. It should be remarked that these and all the pictures obtained with the Kerr camera have been recorded on Super-Pan Press, Type B film; the film records have in all cases been processed for nine (9) minutes in Eastman DK-60 developer at sixty-eight (68) degrees Fahrenheit.

While the operation of the equipment either in the laboratory or in an enclosed gun shed proved entirely satisfactory, it was found impossible to achieve successful synchronization of the unit when used under outdoor conditions on the open ranges. This failure has just recently been traced to the considerable ambient temperature sensitivity of the mercury control thyratron. Excellent synchronization in all cases has since been achieved by the simple expedient of replacing this thyratron by a hydrogen thyratron (type 5C22). A series of tests were recently carried out to determine the accuracy of synchronization of the unit and also its applicability to armor plate investigations. For this purpose, a target consisting of a one-eighth (1/8)-inch sheet of Dural metal was located down range from the gun with its plane at an angle of forty-five (45) degrees to the line of fire. A contact screen consisting of a cardboard sheet on which horizontal conducting bands of aluminum had been sprayed was located approximately eighteen inches behind the target with its plane perpendicular to the line of fire. Connections are made to alternate bands on such a screen so that triggering of the light and camera is effected as soon as the nose of the projectile connects any two bands. Figure 9 shows a picture of a thirty-seven millimeter armor piercing shot which was equipped with a standard, light alloy ballistic cap taken shortly after passage of the projectile through the target plate. It will be noted that the detail (particularly that of the destruction of the ballistic cap) is uniformly good. It will also be observed that triggering has been effected in this instance by the penetration of the inner, armor-piercing core of the shot through the contact screen. It should be noted here that an Ilex

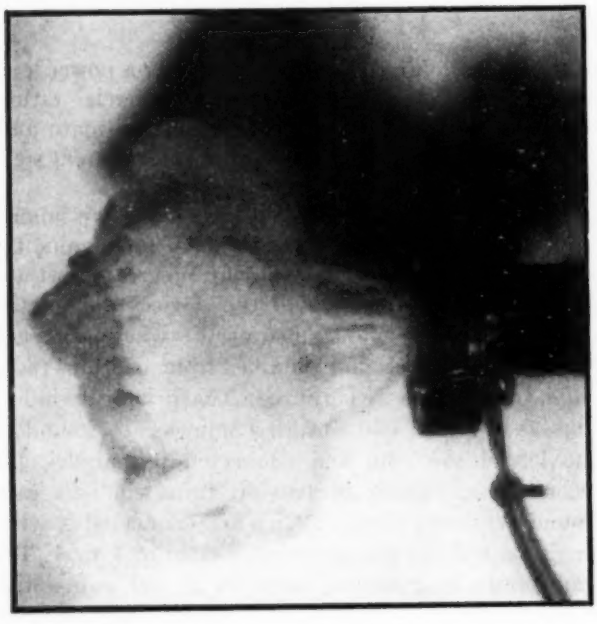


FIG. 8. Kerr cell photograph of gas cloud surrounding 20-mm shell shortly after complete emergence from the muzzle.

FIG. 9. Kerr cell camera photograph of a thirty-seven millimeter shot taken shortly after penetration of a metal target plate.

"Paragon" lens of aperture $F/4.5$ and focal length six point four (6.4) inches equipped with an electromagnetically operated Acme "Synchro" shutter was employed in this last series of tests; the shutter setting was one twenty-fifth (1/25) second and it was operated via an amplifier by an inertia switch mounted on the barrel of the gun.

ACKNOWLEDGMENTS

The authors wish to record their indebtedness to Mr. E. W. Greenwood of this laboratory for his continued assistance in connection with the purely photographic aspects of the problem and for many invaluable suggestions. Our thanks are also due the Defence Research Board of Canada for permission to publish this report.

Effect of a Circular Groundplane on Antenna Radiation*

A. LEITNER
Mathematics Research Group, Washington Square College, New York University, New York, New York

AND

R. D. SPENCE
Michigan State College, East Lansing, Michigan

(Received May 15, 1950)

The field of a quarter-wave-length antenna above a circular conducting disk of zero thickness—the groundplane—is calculated theoretically by use of the wave functions of the oblate spheroid. Assuming a sinusoidal distribution of current on the antenna, we compute the currents on the groundplane, the radiation resistance and the radiation pattern of the system for various values of the radius a. With these results the distortion of antenna radiation by finite groundplanes can be studied. We find good agreement with recent experimental data.

INTRODUCTION

It is well known that the behavior of an antenna placed above a conducting plane of either finite or infinite extent differs from that of an identical antenna in free space. While the effect of an infinite plane can easily be predicted in terms of images, the effect of a plane of finite extent poses a much more difficult problem. The latter problem is of some practical importance since in certain situations it may be desirable to place an actual antenna over a conducting surface of finite extent and since in laboratory tests on scale models of antennas it is often assumed that the effect of the earth may be simulated by a finite conducting plate.

Recently Meier and Summers[1] have shown experimentally that the input impedance of a vertical antenna placed above a finite groundplane depends rather markedly on the size of the plane. The only theoretical attack on this problem which has come to the attention of the present authors is the work of Bardeen[2] on the effect of a circular groundplane on a vertical antenna. In the present paper we reconsider essentially the same problem using a method quite different from that of Bardeen. Our method has the advantage that it is less restricted to planes of small radius although even in our case the computation becomes tedious for planes whose radius is very large compared to the length of the antenna.

In the actual problem considered here the groundplane is assumed to be a perfectly conducting disk of radius a and zero thickness. The antenna is assumed to be one quarter of a wave-length long, infinitely thin, and lies along the axis perpendicular to the disk with its lower end in contact with the disk. The antenna is driven in such a way that the current distribution is sinusoidal. While it is fully realized that the last statement postulates a rather idealized situation it hardly seems worth while to take a more elaborate model for a calculation in which the choice of the antenna is not of primary concern.

THE GREEN'S FUNCTION

The central problem of the present calculation consists of constructing a Green's function from which it is possible to calculate the *total* magnetic field $\mathbf{H}$ at any point in space in terms of the current distribution on the antenna. This means that the Green's function must be so constructed that all the boundary conditions on the electric and magnetic fields $\mathbf{E}$ and $\mathbf{H}$ are automatically satisfied on the groundplane. The Green's function will be proper not only in the present problem, but in any circularly symmetric problem involving the perfectly conducting circular disk. Because of the rather peculiar symmetry of the problem, the magnetic field consists of the single azimuthal component H_φ and therefore the Green's function for the magnetic field has a simple scalar form and not the more complicated dyadic form generally found in electromagnetic problems. A further consequence of this symmetry is the fact that H_φ is

* The research reported in this document has been made possible through partial support and sponsorship extended by the Geophysical Research Directorate of the Air Force Cambridge Research Laboratories under Contract No. AF-19(122)-42. It is published for technical information only and does not represent recommendations or conclusions of the sponsoring agency.

[1] A. S. Meier and W. P. Summers, Proc. I. R. E. **37**, 609 (1949).
[2] J. Bardeen, Phys. Rev. **36**, 1482 (1930).

independent of the azimuthal angle φ. These conditions automatically insure the solenoidal character of **H**. The vector wave equation for **H** reduces to the simple scalar form[3]

$$L[H_\varphi(\rho, z)]=\left(\frac{\partial}{\partial\rho}\frac{1}{\rho}\frac{\partial}{\partial\rho}\rho+\frac{\partial^2}{\partial z^2}+k^2\right)H_\varphi(\rho, z)=0, \quad (1)$$

where ρ and z are the usual cylinder coordinates. These coordinates are not particularly well suited for handling the boundary conditions on the disk and so we introduce a set of oblate spheroidal coordinates η, ξ, φ, which are given by the transformations

$$\rho=a((1-\eta^2)(1+\xi^2))^{\frac{1}{2}}, \quad (2a)$$

$$\varphi=\varphi, \quad (2b)$$

$$z=a\eta\xi, \quad (2c)$$

and for which the surface $\xi=0$ is a disk of zero thickness and radius a. The boundary condition upon the magnetic field on the disk may now be written

$$\partial H_\varphi/\partial\xi=0 \quad \text{when} \quad \xi=0. \quad (3)$$

In the new coordinates Eq. (1) becomes

$$L[H_\varphi(\eta, \xi)]=\left[\frac{\partial}{\partial\eta}(1-\eta^2)\frac{\partial}{\partial\eta}+\frac{\partial}{\partial\xi}(1+\xi^2)\frac{\partial}{\partial\xi}\right.$$

$$\left.-\left(\frac{1}{1-\eta^2}-\frac{1}{1+\xi^2}\right)+k^2a^2(\eta^2+\xi^2)\right]H_\varphi(\eta, \xi)=0. \quad (4)$$

The solutions of (4) constitute a set of orthogonal functions $\psi_{l1}(\eta, \xi)$ which may be written

$$\psi_{l1}(\eta, \xi)=u_{l1}(\eta)v_{l1}(\xi), \quad l=1, 2, 3\cdots. \quad (5)$$

The functions $u_{l1}(\eta)$ and $v_{l1}(\xi)$ are respectively the "angular" and "radial" oblate spheroidal wave functions[4] of the first order.

The Green's function $G(\eta, \xi; \eta', \xi')$ satisfies

$$L[G(\eta, \xi; \eta', \xi')]=-\frac{\delta(\eta-\eta')\delta(\xi-\xi')\delta(\varphi-\varphi')}{h_\eta h_\xi h_\varphi}, \quad (6)$$

where

$$h_\eta=a((\eta^2+\xi^2)/(1-\eta^2))^{\frac{1}{2}}, \quad h_\xi=a((\eta^2+\xi^2)/(1+\xi^2))^{\frac{1}{2}}, \quad (7)$$
$$h_\varphi=a((1-\eta^2)(1+\xi^2))^{\frac{1}{2}},$$

are the metric coefficients of the oblate spheroidal system. Once the Green's function is determined, $H_\varphi(\eta, \xi)$ may be found from

$$H_\varphi(\eta, \xi)=-\int_{\sigma'}[H_\varphi(\eta', \xi')\partial G/\partial n'$$
$$-G\partial H_\varphi(\eta', \xi')/\partial n']d\sigma' \quad (8)$$

in terms of the values of the integrand on the boundary σ'. The sense of $\partial/\partial n'$ is out of the volume enclosed by σ'. The integral is over the surface of the antenna and the groundplane, provided both H_φ and G satisfy the radiation condition at infinity. If we also require that G satisfy the same boundary condition as H_φ,

$$\partial G/\partial\xi=0 \quad \text{when} \quad \xi=0, \quad (9)$$

then the integrand vanishes on the groundplane.

The conditions imposed upon G suffice to determine it uniquely as a solution of (6), and it will be represented in the wave functions (5).

$$G=\frac{1}{2\pi}\sum_{l=1}^{\infty}\frac{u_{l1}(\eta)u_{l1}(\eta')}{N_{l1}}\begin{cases}{}_1v_{l1}(\xi)^{(3)}v_{l1}(\xi'), & \xi\leq\xi',\\ {}_1v_{l1}(\xi')^{(3)}v_{l1}(\xi), & \xi\geq\xi'.\end{cases} \quad (10)$$

This function is symmetric in η, ξ and η', ξ'. The radial functions $v_{l1}(\xi)$ are solutions of the ordinary differential equation

$$\left[\frac{d}{d\xi}(1+\xi^2)\frac{d}{d\xi}+\frac{1}{1+\xi^2}-\alpha_{l1}+k^2a^2(1+\xi^2)\right]v_{l1}(\xi)=0, \quad (11)$$

which one obtains upon separation of variables in (4). The separation constant is constrained to eigenvalues α_{l1}, $l=1, 2, 3, \cdots$, because of conditions of analyticity in the η dimension. Equation (11) has a fundamental set of solutions, ${}^{(1)}v_{l1}$, ${}^{(2)}v_{l1}$. The linear combination

$${}^{(1)}v_{l1}+i^{(2)}v_{l1}={}^{(3)}v_{l1}\underset{\xi\to\infty}{\longrightarrow}-(2q_{l1}/ka)e^{ikr}/kr, \quad (12)$$

called *function of the third kind*, has the property that it behaves like an out-going wave at large ξ, as shown in (12).[5] Thus G obeys the radiation condition as ξ or $\xi'\to\infty$, cf. (10). The factor $(1/2\pi)$ normalizes G in the φ dimension and the norm N_{l1} of u_{l1} does the same for η; the quantity q_{l1} is a factor arising from the normalization of the radial functions, discussed in reference 4. Finally, ${}_1v_{l1}$ is the combination

$${}_1v_{l1}=a_l{}^{(1)}v_{l1}+b_l{}^{(2)}v_{l1} \quad (13)$$

chosen such that the condition (9) be satisfied as ξ or $\xi'\to0$, i.e.

$$a_l{}^{(1)}v_{l1}'(0)+b_l{}^{(2)}v_{l1}'(0)=0. \quad (14)$$

A second equation linear in the a_l, b_l is found when (10) is substituted into (6), the result multiplied by any one of the orthogonal set (u_{l1}) and integrated over the range of η, $-1\leq\eta\leq1$. We obtain the differential equation

$$\left[\frac{d}{d\xi}(1+\xi^2)\frac{d}{d\xi}+\frac{1}{1+\xi^2}-\alpha_{l1}+k^2a^2(1+\xi^2)\right]$$

$$\times\begin{Bmatrix}{}_1v_{l1}(\xi)^{(3)}v_{l1}(\xi')\\ {}_1v_{l1}(\xi')^{(3)}v_{l1}(\xi)\end{Bmatrix}=-\delta(\xi-\xi')/a, \begin{Bmatrix}\xi\leq\xi'\\ \xi\geq\xi'\end{Bmatrix}. \quad (15)$$

[3] Note that neither Eqs. (1) nor (4) represent scalar wave equations.

[4] A comprehensive discussion of these functions in the notation of the present paper may be found in a paper by the present authors. See J. Frank. Inst. 249, 299 (1950).

[5] We assume a dependence $e^{-i\omega t}$ on time.

Integrate this equation with respect to ξ over an interval surrounding the point ξ' and let the interval shrink into this point. The function between braces is continuous but its derivative is by this procedure seen to possess a discontinuity at ξ' of the amount

$$-1/a(1+\xi'^2)={}_1v_{l1}(\xi')^{(3)}v_{l1}'(\xi')-{}_1v_{l1}'(\xi')^{(3)}v_{l1}(\xi'),$$
$$=(ia_l-b_l)W_{l1}({}^{(1)}v_{l1}(\xi'),\ {}^{(2)}v_{l1}(\xi')),$$
$$=(ia_l-b_l)\frac{(-)^{l-1}(2q_{l1}/ka)^2/ka}{1+\xi'^2}, \qquad (16)$$

wherein $W_{l1}({}^{(1)}v_{l1},\ {}^{(2)}v_{l1})$ is the Wronskian of the fundamental set of radial functions whose analytical value is taken from reference 4 and finally substituted above. This leads to a second equation for the constants a_l, b_l and to the result

$${}_1v_{l1}=-\frac{(-)^{l-1}k}{(2q_{l1}/ka)^2}[{}^{(2)}v_{l1}'(0)^{(1)}v_{l1}-{}^{(1)}v_{l1}'(0)^{(2)}v_{l1}]. \qquad (17)$$

Let us call the function inside brackets ${}_2v_{l1}$. Then

$$G=\frac{k}{2\pi}\sum_{l=1}^{\infty}\frac{(-)^l u_{l1}(\eta)u_{l1}(\eta')}{N_{l1}(2q_{l1}/ka)^{2(3)}v_{l1}'(0)}$$
$$\times\begin{cases}{}_2v_{l1}(\xi)^{(3)}v_{l1}(\xi'),\ \xi\leq\xi',\\ {}_2v_{l1}(\xi')^{(3)}v_{l1}(\xi),\ \xi\geq\xi'.\end{cases} \qquad (18)$$

On the groundplane $\xi=0$. It can be shown that

$$(-)^{l-1}{}_2v_{l1}(0)/(2q_{l1}/ka)^2=1/ka. \qquad (19)$$

Therefore

$$G(\eta,0;\eta',\xi')$$
$$=-\frac{1}{2\pi a}\sum_{l=1}^{\infty}u_{l1}(\eta)u_{l1}(\eta')^{(3)}v_{l1}(\xi')/N_{l1}{}^{(3)}v_{l1}'(0). \qquad (20)$$

At large distances $\xi\longrightarrow\infty$, and

$$G(\eta,\infty;\eta',\xi')$$
$$=\frac{ka}{4\pi}\frac{e^{ikr}}{r}\sum_{l=1}^{\infty}\frac{(-)^{l-1}u_{l1}(\eta)u_{l1}(\eta')_2v_{l1}(\xi')}{N_{l1}q_{l1}{}^{(3)}v_{l1}'(0)}. \qquad (21)$$

DERIVATION OF THE MAGNETIC FIELD

The integral (8) for H_φ is taken over the antenna, a section of the z-axis where $\eta'=1$ and $0\leq\xi'\leq\pi/2ka$, $\partial/\partial n'=\partial/h_{\eta'}\partial\eta'$ and $d\sigma'=(h_{\xi'}h_{\varphi'})_{\eta'=1}d\xi'd\varphi'$. We get

$$H_\varphi=-2\pi a\int_0^{\pi/2ka}[(H_{\varphi'}\partial G/\partial\eta'$$
$$-G\partial H_{\varphi'}/\partial\eta')(1-\eta'^2)]_{\eta'=1}d\xi'. \qquad (22)$$

On the surface of the antenna $H_{\varphi'}$ approaches

$$I_0\cos kz'/2\pi\rho'=I_0\cos ka\xi'/2\pi a(1-\eta^{\frac{1}{2}})^{\frac{1}{2}}(1+\xi'^2)^{\frac{1}{2}},$$

where $I_0\cos kz'$ is the current on the antenna, and $\partial H_{\varphi'}/\partial\eta'$ approaches

$$I_0\cos ka\xi'/2\pi a(1-\eta'^2)^{\frac{1}{2}}(1+\xi'^2)^{\frac{1}{2}}.$$

When these values and Eqs. (20) and (21) are substituted into the integral one finds

$$H_\varphi(\eta,0)=-\frac{I_0}{\pi a}\sum_{l=1}^{\infty}{}^{(3)}g_l u_{l1}(\eta)/N_{l1}{}^{(3)}v_{l1}'(0), \qquad (23)$$

$$H_\varphi(\eta,\infty)=\frac{kaI_0}{2\pi}\frac{e^{ikr}}{r}$$
$$\times\sum_{l=1}^{\infty}(-)^{l-1}{}_2g_l u_{l1}(\eta)/N_{l1}q_{l1}{}^{(3)}v_{l1}'(0). \qquad (24)$$

In both expressions

$$g_l=\int_0^{\pi/2ka}[v_{l1}(\xi')\cos(ka\xi')/(1+\xi'^2)^{\frac{1}{2}}]d\xi'. \qquad (25)$$

The different sub- and superscripts on g_l in (23) and (24) refer to the type of radial function implied in the integrals. These can be evaluated explicitly. We define

$$w_{l1}=(1+\xi^2)^{\frac{1}{2}}v_{l1}, \qquad (26)$$

and find on substitution into (11) that these functions satisfy

$$w_{l1}''+k^2a^2w_{l1}=\alpha_{l1}w_{l1}/(1+\xi^2). \qquad (27)$$

Let

$$y=\cos ka\xi. \qquad (28)$$

Then

$$\left|yw_{l1}'-y'w_{l1}\right|_{\xi_1}^{\xi_2}=\alpha_{l1}\int_{\xi_1}^{\xi_2}[yw_{l1}/(1+\xi^2)]d\xi, \qquad (29)$$

and

$${}^{(3)}g=-{}^{(3)}v_{l1}'(0)[1-ka^{(3)}w_{l1}(\pi/2ka)/{}^{(3)}w_{l1}'(0)]/\alpha_{l1}, \qquad (30)$$

$${}_2g=ka_2w_{l1}(\pi/2ka)/\alpha_{l1}. \qquad (31)$$

Finally

$$H_\varphi(\eta,0)=\frac{I_0}{\pi a}\sum_{l=1}^{\infty}A_l u_{l1}(\eta), \qquad (32)$$

$$H_\varphi(\eta,\infty)=\frac{I_0k^2a^2}{2\pi}\frac{e^{ikr}}{r}\sum_{l=1}^{\infty}B_l u_{l1}(\eta), \qquad (33)$$

where

$$A_l=\frac{1-ka^{(3)}w_{l1}(\pi/2ka)/{}^{(3)}w_{l1}'(0)}{\alpha_{l1}N_{l1}}, \qquad (34)$$

$$B_l=(-1)^{l-1}{}_2w_{l1}(\pi/2ka)/\alpha_{l1}N_{l1}q_{l1}{}^{(3)}v_{l1}'(0). \qquad (35)$$

DISTRIBUTIONS OF CURRENT ON THE GROUNDPLANE

The eigenfunction expansion (32) for the magnetic field on the groundplane is analytic term by term but

the coefficients A_l diverge at least as slowly as the sequence of integers l. We will omit the proof of this assertion here.

The divergence is due to singularities in H_φ which occur when $|\eta|=1$, that is, when $\rho=a(1-\eta^2)^{\frac{1}{2}}=0$. These are the points at the centers of both sides of the disk. Now the part of H_φ which is due to the antenna alone has the singularity

$$I_0/4\pi\rho=I_0/4\pi a(1-\eta^2)^{\frac{1}{2}} \text{ as } \rho\to 0.^{6} \tag{36}$$

The remaining part of H_φ represents the field scattered by the disk. We will show that it must have a singularity of the form

$$I_0\eta/4\pi a(1-\eta^2)^{\frac{1}{2}} \text{ as } \rho\to 0 \tag{37}$$

on the surface of the disk. The behavior of H_φ would now be

$$H_\varphi(\eta, 0)\to I_0(1+\eta)/4\pi a(1-\eta^2)^{\frac{1}{2}}, \text{ as } \rho\to 0. \tag{38}$$

This will insure continuity between the radial current flowing into the center on the disk and the current flowing upward along the antenna. The latter is $I_0 \cos kz$ and has the value I_0 at the base. The current on the disk (flowing into the center $\eta=+1$ on the top-, and out of the center $\eta=-1$ on the underside) is simply $2\pi\rho H_\varphi(\eta, 0)$. It is clear from Eq. (38) that at $\eta=+1$ the groundplane current will be I_0 and will vanish at $\eta=-1$.

Note also that the scattered part of H_φ is now odd in η, i.e., odd about the plane of the disk. This is a general property in problems involving plane scatterers subject to the boundary condition (3) and can be proven independently.[7]

The even part of the expansion (32) for $H_\varphi(\eta, 0)$ are the terms for which the $(l-1)$ are even. This part corresponds to the value of the magnetic field of the antenna alone on the surface of the disk. There exist explicit expressions for this quantity as already mentioned.[6] The divergence in the even terms is therefore removed when they are replaced by these expressions.

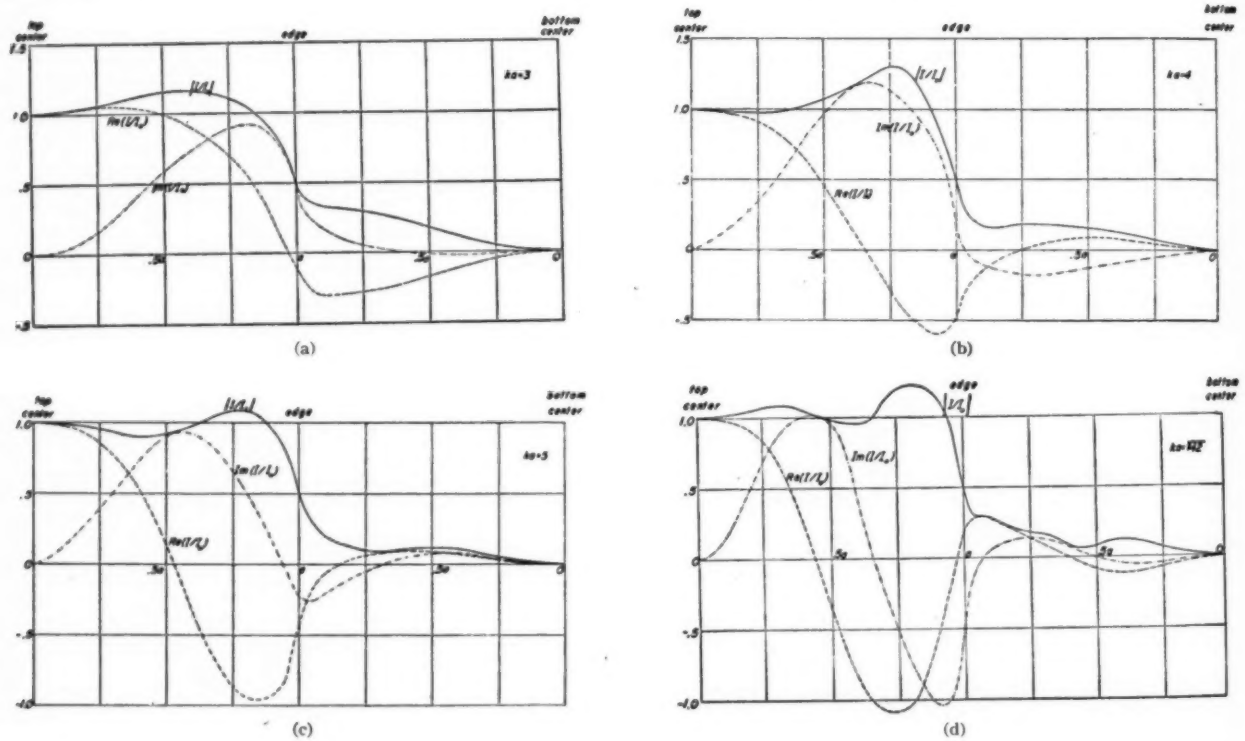
The divergence in the odd part is removed by expanding the singularity in (37) in the odd angular functions,

$$\eta/(1-\eta^2)^{\frac{1}{2}}=\sum_{l=2}^{\infty}{}' C_l u_{l1}(\eta). \tag{39}$$

The series is over alternate integers l (indicated by a prime on the summation sign), and the sequence of coefficients

$$C_l=(1/N_{l1})\int_{-1}^{1}[\eta u_{l1}(\eta)/(1-\eta^2)^{\frac{1}{2}}]d\eta, \ (l-1) \text{ odd}, \tag{40}$$

must necessarily diverge.



FIGS. 1(a) to (d). Ratio of radial current I (on both sides of the groundplane) to the amplitude I_0 of the current on the antenna. Plotted against distance from the center for the cases when $ka=3, 4, 5, (42)^{\frac{1}{2}}$.

[6] One may obtain this result upon substitution into the expressions for the field of our antenna which are given in the literature. See for instance J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill Book Company, Inc., New York, 1941), p. 457, Eq. (75). (Practical units are used in the present paper.)

[7] See, for example, A. Leitner, J. Acous. Soc. Am. **21**, 331 (1949).

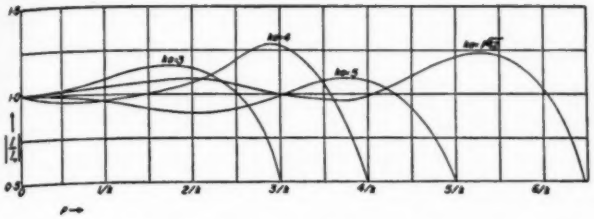


FIG. 2. Absolute values of the ratio of radial current I (on the *antenna* side of the groundplane) to the amplitude I_0 of the current on the antenna. Plotted to common scale against distance from the center for the cases when $ka=3, 4, 5, (42)^{\frac{1}{2}}$.

Finally, the current $I(\eta)$ on the groundplane has the value

$$I(\eta)=2\pi\rho H_\varphi(\eta,0)$$
$$=(I_0/2)[-i\exp[ik(\rho^2+(\lambda/4)^2)^{\frac{1}{2}}]$$
$$+\eta+(1-\eta^2)^{\frac{1}{2}}\cdot\sum_{l=2}^{\infty}{}'(4A_l-C_l)u_{l1}(\eta)]. \quad (41)$$

Numerical computations show that the remaining series in this expression converges.

The ratios of $I(\eta)$ to I_0 are plotted against distance from the base of the antenna in Fig. 1. Since $I(\eta)$ is continuous at the edge $\rho=a$, it was possible to plot the values on both sides continuously on one graph.

As ka increases, the currents on the underside become smaller, while on the antenna side the average of the absolute value becomes more nearly unity for each ka. Figure 2 illustrates this latter point more clearly. Theer we have plotted the currents on the *antenna* side on one graph for the various values of the parameter ka, with a common scale for the distance from the center. The larger ka correspond to larger groundplanes.

The trends revealed in the figures are consistent with the case of an infinite groundplane. In that case no current would flow on the underside, while the absolute value would be I_0 at all points on the upper side.

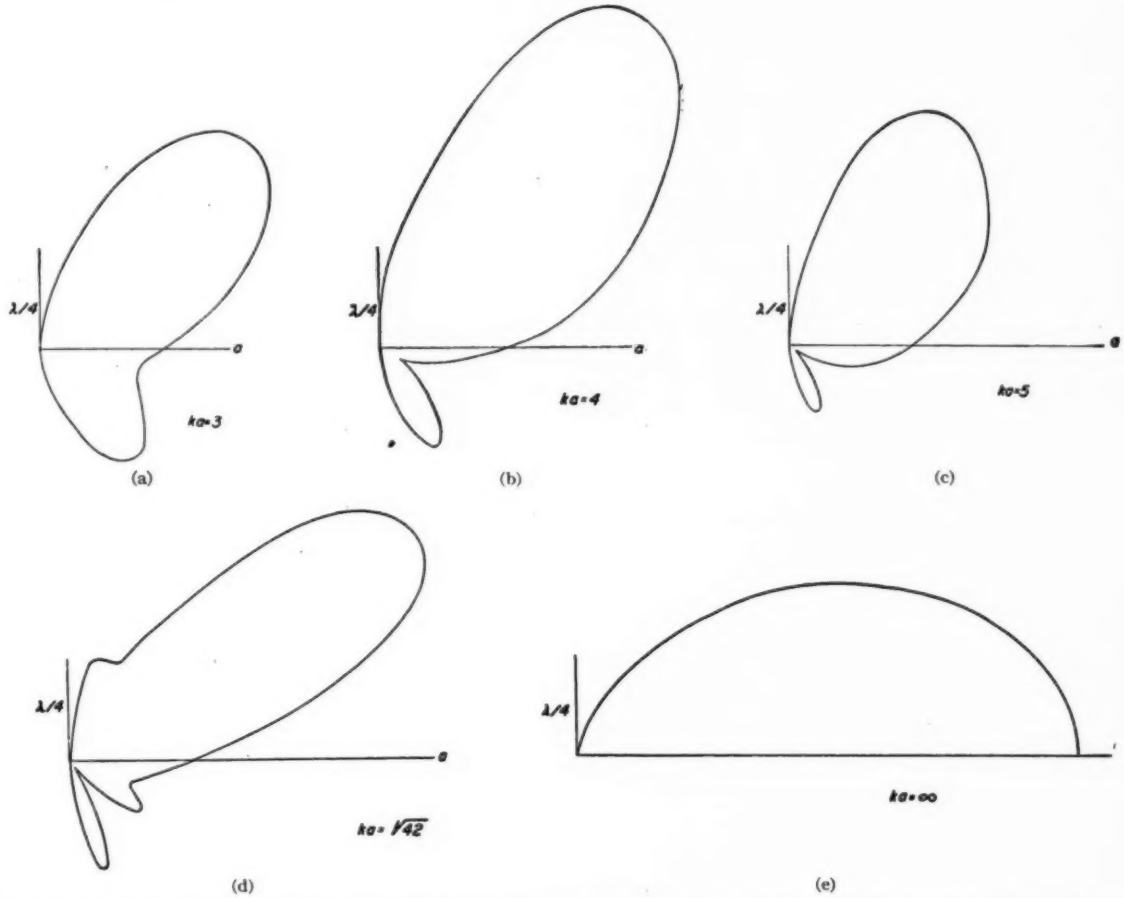
RADIATION PATTERNS

We next study the field at large distances from the radiating system.

As $r\to\infty$ the electric vector is entirely transverse to the radial direction,

$$E\to E_\eta\to(\mu_0/\epsilon_0)^{\frac{1}{2}}H_\varphi, \quad (42)$$

where μ_0 and ϵ_0 are the free space permeability and dielectric constant, respectively. So the Poynting flux



FIGS. 3(a) to (e). Radiation patterns (polar graphs in any azimuthal plane) of a quarter wave-length antenna situated above circular groundplanes for the cases when $ka=3, 4, 5, (42)^{\frac{1}{2}}, \infty$.

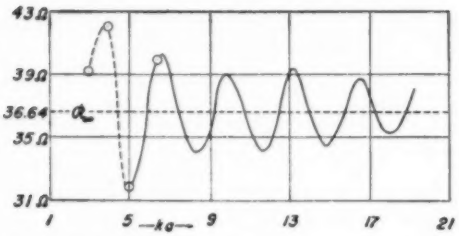


FIG. 4. Radiation resistance of the system of antenna and groundplane plotted against ka. *Solid curve:* experimental data of Meier and Summers. *Dotted curve:* extrapolation of data. *Circles:* calculated values. *Dotted horizontal line:* corresponds to $ka=\infty$.

becomes

$$S=\tfrac{1}{2}E_\eta\tilde{H}_\varphi$$

$$=(I_0^2/8\pi r^2)(\mu_0/\epsilon_0)^{\frac{1}{2}}(ka)^4\Big|\sum_{l=1}^{\infty}B_l u_{l1}(\eta)\Big|^2. \quad (43)$$

This quantity is to be compared to the case of the infinite groundplane, when the Poynting flux is the same as for the half-wave-length antenna in the region $\theta\leq\pi/2$, and vanishes below. In this familiar case[8]

$$S_\infty=(I_0^2/8\pi r^2)(\mu_0/\epsilon_0)^{\frac{1}{2}}\cos^2[\pi(\cos\theta)/2]/\sin^2\theta.^9 \quad (44)$$

(The subscript ∞ refers to the infinite groundplane.)

A polar graph of the dependence of S on the angle θ is the radiation pattern. It is drawn, to common scale, for each of the parameter values $ka=3, 4, 5, (42)^{\frac{1}{2}}$ and ∞, in Fig. 3.

We see that the patterns 3(a) to (d) differ greatly from the pattern 3(e). The radiation field shows no sign of approaching the case corresponding to the isolated half-wave-length antenna. Technical applications may possibly be found for the peculiar directivity of the finite groundplane system illustrated by the figure: while the isolated antenna radiates its energy most strongly in a direction perpendicular to the antenna, the finite groundplane system radiates preponderantly in an oblique direction.

[8] See reference 6, p. 441, Eq. (11).
[9] Note that $\eta\to\cos\theta$ as $r\to\infty$.

RADIATION RESISTANCE

The radiation resistance of the finite groundplane system is calculated from the Poynting flux, and becomes

$$\mathfrak{R}=(1/2\pi)(\mu_0/\epsilon_0)^{\frac{1}{2}}(ka)^4\sum_{l=1}^{\infty}N_{l1}|B_l|^2 \text{ ohms}, \quad (45)$$

while in the infinite case

$$\mathfrak{R}_\infty=36.64 \text{ ohms}, \quad (46)$$

exactly one-half the resistance of the half-wave antenna.

These values are plotted in Fig. 4. The solid line represents the experimental data of Meier and Summers, taken on thin wire antennas approximately one quarter of a wave-length long with thin circular conducting plates as groundplanes. The results of our calculation are entered as circles and lie on the experimental curve or an extrapolation of this curve, drawn as a dotted line.

The value of $\mathfrak{R}$ never differs greatly from $\mathfrak{R}_\infty$ and, as a function of ka, oscillates about $\mathfrak{R}_\infty$, approaching this value asymptotically, as is to be expected when $ka\to\infty$.

When one wishes to measure the resistance of an isolated antenna, the rule has been to mount the test model over a groundplane, and consider the resistance of the system as the resistance of the isolated antenna. The error in this measurement depends on the relative size of the groundplane, as follows clearly from the curve in Fig. 4. An improvement over this procedure suggests itself from the nature of the curve. Rather than a single groundplane, one might use a whole series of them, similar in shape but differing in size, for each single antenna to be tested. A graph of resistance against the size of the groundplane will resemble the curve in Fig. 4. From such a curve it should be possible to estimate the position of the asymptote $\mathfrak{R}=\mathfrak{R}_\infty$ with greater accuracy.

We wish to thank Professor C. Kirkuchi for reading the manuscript and are particularly grateful to Miss J. Bruno for performing the computations.

Turbulence in Apparatus for Measurement of Streaming Double Refraction

H. G. Jerrard
Physics Department, University College, Southampton, England
(Received March 16, 1950)

The different apparatus used by some thirty investigators, who have induced double refraction in a liquid contained in the annular gap between two concentric cylinders, are examined with particular reference to the onset of turbulence. In the case where the outer cylinder alone rotates, curves are given which easily enable the critical gradients corresponding to the dimensions of the apparatus used to be found. The results of the examination are so tabulated that present and future workers in this field can quickly decide on the significance of the measurements obtained by any of the investigators. An explanation is put forward of the surprising fact that with either the inner or outer cylinder rotating measurements of double refraction apparently show no very sharp break between regions of laminar and turbulent flow. The value of making simultaneous measurements of streaming double refraction and viscosity is pointed out and the advantages of using an apparatus with rotating outer and stationary inner cylinder are briefly summarized. Mention is made of optical design.

I. INTRODUCTION

It was observed by Maxwell[1] that Canada balsam, although normally isotropic, becomes birefringent when subjected to a shearing stress. The name streaming double refraction is usually applied to this effect in analogy to the terms magnetic and electrical double refraction. Since the original work of Maxwell, streaming double refraction in liquids has been extensively studied and is now recognized as a powerful method for the determination of molecular size and shape.[2]

The essential condition for the appearance of streaming double refraction is the production of a velocity gradient in the moving liquid, which is then examined by a suitable optical method. As a prelude to the construction of apparatus for streaming double refraction experiments, the descriptions of the apparatus used by previous workers in this field have been examined. It appears that only two methods are in use to produce the required velocity gradient. One method is to produce a gradient by flow of a liquid in a narrow tube of rectangular cross section and the other by motion of a liquid confined in the annular gap between two concentric cylinders, one of which is stationary, the other rotating.

The first method is suitable for the production of gradients of the order of 2000–3000 sec.$^{-1}$ but which, however, are not constant across the tube. For pure liquids the gradient varies linearly from zero at the center to a high value near the walls; this corresponds to a parabolic distribution of velocity across the tube. Colloidal solutions on the other hand show a non-linear variation of gradient from zero at the center to a maximum at the walls and give a non-parabolic distribution of velocity.[3]

In both cases, therefore, there is no uniquely defined value of the velocity gradient which can be correlated with a given double refraction measurement. The second method is capable not only of producing an almost constant velocity gradient across the gap for both homogeneous and colloidal solutions but enables very high gradients (of the order of 30,000 sec.$^{-1}$) to be obtained.

The majority of observers have used this second method but unfortunately many workers have paid insufficient attention to the possibility of turbulence. The latter is, however, of prime importance since streaming double refraction is not amenable to theoretical treatment in the turbulent region and the results obtained for the double refraction, although of the right order of magnitude, can have no quantitative significance. It is a surprising fact, however, that except in a few isolated cases of colloidal solutions the onset of turbulence does not seem to produce any sharp break in curves showing the variation of double refraction with velocity gradient.[4]

In this paper the results of the examination, from the point of view of turbulence, of the concentric cylinder apparatus described by various authors are presented and reasons why the onset of turbulence may escape notice are given. It is hoped that the results may be of use to other workers in the fields of streaming double refraction and viscometry.

II. THEORY

In the concentric cylinder method, either the inner cylinder can rotate with the outer stationary or vice versa. The equations of motion for steady flow are the same in both cases but there is considerable difference in the velocity at which turbulence first occurs, because the distribution of centrifugal forces for the case of the rotating outer cylinder is such as to stabilize the motion of the fluid, whereas with the rotating inner cylinder this is not so.

1. Steady Motion

Suppose V be the velocity at any point P of an incompressible viscous fluid in steady motion between

[1] J. C. Maxwell, *Scientific Papers* (Cambridge University Press, London, 1890), Vol. II, p. 379. Proc. Roy. Soc. **A22**, 46 (1873).
[2] See review by J. T. Edsall, *Advances in Colloid Science* (Interscience Publishers, Inc., New York, 1942), Vol. 1, 269.
[3] A. S. C. Lawrence, Proc. Roy. Soc. **A148**, 59 (1935); W. Philippoff, Kolloid. Zeits. **75**, 142 (1936).
[4] C. Sadron, J. Phys. Radium **7**, 263 (1936); Schweiz. Arch. angew. Wiss. Tech. (3–4) **158**, 8 (1937).

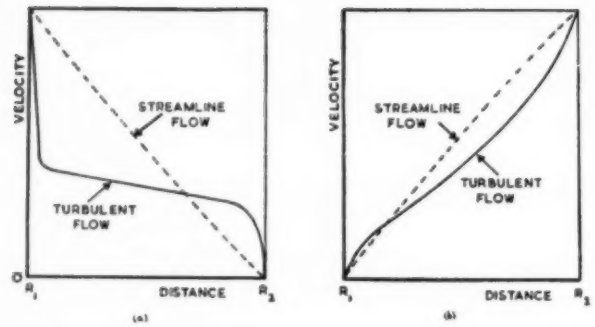


FIG. 1. Velocity distribution across the annular gap. (a) Inner cylinder rotating (after Wattendorf). (b) Outer cylinder rotating (after G. I. Taylor).

two infinitely long concentric rotating cylinders of radii R_1 and R_2, $(R_2>R_1)$. If r is the distance of P from the axis and if Ω_1 and Ω_2 are the angular velocities of the cylinders, then from the Navier-Stokes equations,[5] it may be shown that:

$$V=Ar+(B/r),$$

where

$$A=\frac{R_2^2\Omega_2-R_1^2\Omega_1}{R_2^2-R_1^2} \quad \text{and} \quad B=\frac{R_1^2R_2^2(\Omega_1-\Omega_2)}{R_2^2-R_1^2}.$$

The linear velocity gradient G is given by

$$G=dV/dr=A-(B/r^2)$$

and the change of gradient with distance is

$$dG/dr=2B/r^3.$$

If either cylinder is stationary and the other has an angular velocity Ω, then for small gap widths, the change in gradient across the gap is given by $dG=\Omega$, and the average gradient by $G=\Omega R/s$. In the latter $R=R_2\doteq R_1$, $s=R_2-R_1$ and $s\ll R$. Hence, $dG/G=s/R$. This ratio may be taken as representative of the constancy of the velocity gradient: it will be called the constancy factor.[6] Clearly if s/R is small, then so also is the error in assuming the gradient to have the constant value $(R\Omega)/s$ across the gap. The importance of this will be seen later.

2. Turbulent Motion

The stability for symmetrical disturbances of a viscous liquid in steady motion between concentric cylinders has been the subject of a very thorough theoretical and experimental treatment by Taylor.[7]

(a) *Inner cylinder rotating.* The motion of the liquid in this case is two-dimensional and according to Taylor is stable for angular velocities of the cylinder not exceeding a critical value Ω_c, given by the equation

$$\Omega_c^2=\frac{\pi^4\nu^2(R_1+R_2)}{2Ps^3R_1^2}, \tag{1}$$

which corresponds to a critical Reynolds number Re such that

$$\mathrm{Re}=\frac{(R_1\Omega_c)s}{\nu}=\pi^2\left(\frac{R_1+R_2}{2Ps}\right)^{\frac{1}{2}}. \tag{2}$$

In Eqs. (1) and (2), ν is the kinematic viscosity, s, R_1 and R_2 have the same meaning as before and P is a constant which Taylor gives as

$$P=0.0571[1-0.652(s/R_1)]+0.00056[1-0.652(s/R_1)]^{-1}.$$

When s/R_1 is small, P may be taken as 0.057. If s/R_1 is not small the value of P above, which neglects second-order terms of s/R_1 and Taylor states is to an order of approximation of about 1 percent, holds when $s/R_1\gtrsim 0.33$. Some other limitations of Taylor's criteria are mentioned later. The validity of Eq. (1) was investigated experimentally by Lewis,[8] who states that it holds for values of s/R_1 as high as 0.71 and for liquids whose coefficients of kinematic viscosity vary from 0.006 to 0.018 stokes. A less detailed theoretical study, which has the merit that it involves no approximations based on the thinness of the liquid layer, has been given by Synge.[9]

As soon as the velocity given by Eq. (1) is exceeded the laminar motion is succeeded by a three-dimensional motion such that the liquid breaks up into symmetrical ring-shaped vortices spread at regular intervals along the length of the cylinder: these vortices rotate alternately in opposite directions. It is true that this is not exactly turbulence, as we generally understand it, since the pattern is regular and stationary, but it is an intermediate stage between laminar and turbulent motion.

In an experimental investigation of the velocity

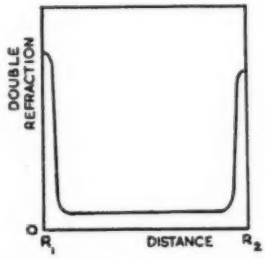


FIG. 2. Distribution of double refraction across annular gap for case of rotating inner cylinder.

[5] See for example H. Lamb, *Hydrodynamics* (Cambridge University Press, London, 1930), 5th Ed. Chapter 11.

[6] For the dimensions of the apparatus given in Tables I and II, the radii of the outer and inner cylinders are not always approximately equal. In order that the values of the constancy factor shall be comparable, the inner cylinder radius, R_1, has been used instead of R.

[7] G. I. Taylor, Phil. Trans. **A223**, 289 (1923); G. I. Taylor, Proc. Roy. Soc. **A102**, 541 (1923); G. I. Taylor, Proc. Roy. Soc. **A157**, 546 (1936).

[8] J. W. Lewis, Proc. Roy. Soc. **A117**, 388 (1928). For a description of the apparatus used by Lewis see J. W. Lewis and E. N. da C. Andrade, J. Sci. Inst. **1**, 373 (1923).

[9] J. L. Synge, Proc. Roy. Soc. **A167**, 250 (1938).

distribution in the gap, Taylor[10] showed that there was a large change in velocity near the walls but throughout some 83 percent of the gap the product of linear velocity and radius was constant. This confirmed the work of Wattendorf[11] who obtained a velocity distribution curve of the type shown in Fig. 1a. These facts can be explained on the theory that the dynamics of turbulent motion should be regarded as a transport of vorticity[12] and is important because it explains the phenomenon mentioned by Snellman[13] and Buchheim, Stuart, and Menz[14] that in some cases, near the surfaces of the cylinders doubly refracting layers have been seen and between them a domain without any double refraction.[15] Thus if V_1 and V_2 be the velocities at points of the liquid

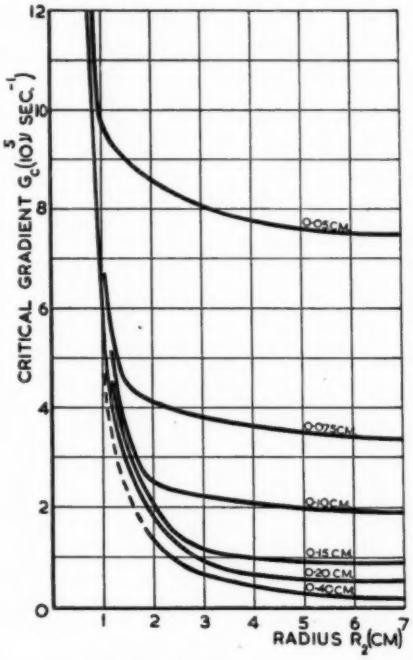


FIG. 3. Variation of critical velocity gradient with radius of outer cylinder. Gap widths 0.05, 0.075, 0.10, 0.15, 0.20, 0.40 cm. Outer cylinder rotating.

distance r and $r+x$, respectively from the axis, then

$$V_1 r = V_2(r+x) = K \text{ (a constant)},$$

so that

$$V_2 - V_1 = Kx/r(r+x) \qquad (3)$$

or the linear velocity gradient, $(V_2-V_1)/x = K/(r(r+x))$. The gradient, therefore, is approximately inversely proportional to r^2 and will be small at the center of the gap where Eq. (3) holds and large near the surfaces of

[10] G. I. Taylor, Proc. Roy. Soc. **A151**, 494 (1935).
[11] F. L. Wattendorf, Proc. Roy. Soc. **A148**, 565 (1935).
[12] G. I. Taylor, Proc. Roy. Soc. **A135**, 685 (1932).
[13] O. Snellman, Ark. Kemi. Min. Geol. **19A**, 30, 3 (1945).
[14] Buchheim, Stuart, and Menz, Zeits. f. Physik **112** (7–8), 407 (1939).
[15] In a private communication, Dr. Snellman states his was an experimental observation with a rotating inner and stationary outer cylinder.

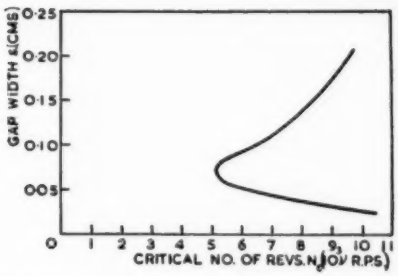


FIG. 4. Variation of critical number of r.p.s. with width of gap. Outer cylinder radius = 1.25 cm.

the cylinders where the velocity change is considerable (see Fig. 1a). Doubly refracting layers will, therefore, be more easily seen at the cylinder surfaces, the distribution of double refraction across the gap being as in Fig. 2. If a well-aligned optical system is used for observation such a distribution may be detected,[16] but with a less efficient system the double refraction may appear as a mean value, uniform across the gap. The mean value may be greater than that which would exist for laminar flow if the layers near the cylinder surfaces are traversed by the light and smaller (even zero) if the middle region is traversed. Thus as previously mentioned the onset of turbulence may cause no sharp break in the curve showing variation of double refraction with gradient but perhaps a slight departure from the curve for laminar flow to give either too high or too low values of the double refraction. This has been detected by Sadron.[17]

(b) *Outer cylinder rotating.* In this case the critical velocity is very much higher than that when the inner cylinder is the rotor. Taylor[18] gives the value of Ω_c in the form of a curve which shows $\log_{10}(\Omega R_2 s/\nu)$ plotted against $\log_{10}(s/R_2)$. This curve is good for Taylor's purpose, but is somewhat tedious to use if one wishes to find quickly the critical speeds and gradients for a specific piece of apparatus. For this reason curves shown in Fig. 3 have been drawn from data derived from Fig. 11 of Taylor's article: they show the values of the critical gradients plotted against the radii of the outer cylinder for various gap widths. Figure 3 must be understood as giving a limit to the critical gradient below which no turbulence occurs. Because of the inevitable inaccuracy in obtaining data from Taylor's published diagram, Fig. 3 does not give with equal certainty the upper limits. These curves should prove of value in the design and use of high speed concentric cylinder viscometers.

From these curves it is easy to calculate the critical number of revolutions per sec. N_c for any system (i.e. the number of revolutions per sec. above which turbulence occurs) and a graph of s against N_c for a fixed value of the radius takes the form shown in Fig. 4. This

[16] In the experiments of Sadron on pure liquids and colloidal solutions (Schweiz. Arch. angew. Wiss. Tech. (3–4) **158**, 8 (1937–(38)) this appears as a non-uniform illumination of the half-shadow.
[17] C. Sadron, J. Phys. Radium **7**, 263 (1936).
[18] G. I. Taylor, Proc. Roy. Soc. **A157**, 546 (1936).

TABLE I. Dimensions of cylinders in apparatus used by various investigators and critical velocity gradients—inner cylinder rotating.

Author	Inner cylinder rad. (R_1) cm	Gap width (s) cm	Gradient constancy factor $\frac{s}{R_1}\frac{dG}{G}=-(10^{-2})$	Critical angular velocity Ω_c (ν-rad. sec.$^{-1}$)	Critical velocity gradient G_c (ν sec.$^{-1}$)	Critical No. of revs. N_c (ν r.p.s.)
Kundt[a] (a)	0.95	0.25	26.3	393	1493	63
Kundt[a] (b)	2.50	0.40	16.0	113	706	18
Kundt[a] (c)	0.95	0.10	10.5	1420	13490	226
de Metz[b] (a)	1.75	0.50	28.6	104	363	16
de Metz[b] (b)	1.75	1.00	57.1	44	77	7
Umlauf[c] (a)	2.33	0.37	15.9	131	825	21
Umlauf[c] (b)	2.10	0.60	28.6	72	252	11
Krueger[d]	2.00	0.25	12.5	250	2000	40
Vorlander and Walter[e]	1.05	0.10	9.5	1340	14070	213
Vorlander and Kirchner[f]	1.10	0.10	9.1	1330	14630	212
Signer[g, h]	2.300	0.195	8.5	330	3897	52
Boeder[i] (a)	4.05	0.20	4.9	235	4755	37
Boeder[i] (b)	4.05	0.10	2.5	655	26520	104
Vorlander and Fisher[j, k]	9.20	0.05	0.54	1215	223600	193
Signer and Gross[l]	2.4700	0.0252	1.0	6576	644500	1047
Sadron[m, n*] (a)	2.498	0.052	2.1	2220	106600	353
Sadron[m, n*] (b)	2.420	0.130	5.4	581	10680	91
Sadron[m, n*] (c)	2.500	0.050	2.0	2350	117500	374
Sadron[m, n*] (d)	2.528	0.022	0.87	7960	914700	1267
Nitschmann[o]	2.475	0.0427	1.7	2990	173300	476
Buchheim,† Stuart and Menz[p] (a)	4.991	0.039	0.78	2400	307100	382
Buchheim,† Stuart and Menz[p] (b)	4.871	0.159	3.3	299	9160	47
Buchheim,† Stuart and Menz[p] (c)	4.770	0.260	5.45	146	2678	23
Nitschmann and Guggisberg[q]	2.4800	0.0209	0.84	8680	1030000	1381
Snellman and Bjornstahl[r, s] (a)	3.50	0.05	1.43	1981	138700	315
Snellman and Bjornstahl[r, s] (b)	3.45	0.10	2.90	711	24550	113
Snellman and Bjornstahl[r, s] (c)	3.40	0.15	4.41	393	8910	63
de Rosset[t]	5.266	0.0268	0.51	4094	804400	652
Edsall[u, v] *et al.*	2.512	0.025	0.99$_5$	6598	663000	1050
Tsvetkov[w] *et al.*	Data published insufficient to enable critical values to be calculated.					

[a] A. Kundt, Wied. Ann. **13**, 110 (1881).
[b] G. de Metz, Ann. Phys. (Leipzig) **35**, 497 (1888).
[c] K. Umlauf, Ann. Phys. (Leipzig) **45**, 304 (1892).
[d] E. Kreuger, Zeits. f. physik. Chemie **A109**, 438 (1924).
[e] D. Vorländer and R. Walter, Zeits. f. physik. Chemie **A118**, 1 (1925).
[f] D. Vorländer and U. Kirchner, Zeits. f. physik. Chemie **A152**, 47 (1930).
[g] R. Signer, Zeits. f. physik. Chemie **A150**, 257 (1930).
[h] This apparatus was also used by G. Boehm and R. Signer, Helv. Chim. Acta. **14**, 1370 (1931).
[i] P. Boeder, Zeits. f. Physik **75** (3–4), 258 (1932).
[j] D. Vorländer and J. Fischer, Berichte **65B**, 1756 (1932).
[k] This apparatus was also used by D. Vorländer and P. Specht, Zeits. f. physik. Chemie **A178**, 93 (1936).
[l] R. Signer and H. Gross, Zeits. f. physik. Chemie **A165**, 161 (1933).
[m] C. Sadron, C.R. Acad. Sci. (Paris) **202**, 404 (1936); Schweiz. Arch. angew. Wiss. Tech. (3–4) **158**, 8 (1937); J. Phys. Radium **7**, 263 (1936).
[n] See also C. Sadron and H. Mosimann, J. Phys. Radium. **9**, 384 (1938); Sadron, Bonot, and Mosimann, J. Chim. Phys. **36**, 78 (1939).
[*] Although Taylor's formula as quoted by Sadron is incorrect it seems that this is a typographical error as the correct formula appears to have been used to calculate Ω_c and G_c.
[o] H. Nitschmann, Helv. Chim. Acta **21**, 315 (1938).
† These authors appear to have made a numerical mistake in copying Taylor's formula. Their published values of N_c must be divided by $\sqrt{2}$.
[p] Buchheim, Stuart, and Menz, Zeits. f. Physik **112** (7–8), 407 (1939).
[q] H. Nitschmann and H. Guggisberg, Helv. Chim. Acta **24**, 434, 574 (1941).
[r] O. Snellman and Y. Bjornstahl, Kolloid. Beih. **52**, 403 (1941).
[s] This apparatus was also used by O. Snellman, Ark. Kemi. Min. Geol. **B19** (No. 5), 1 (1944); Acta. Chem. Scand. **1**, 291 (1947).
[t] A. J. de Rossett, J. Chem. Phys. **9**, 766 (1941).
[u] Edsall, Gordon, Mehl, Scheinberg, and Mann, Rev. Sci. Inst. **15**, 243 (1944).
[v] This apparatus was also used by J. F. Foster and J. T. Edsall, J. Am. Chem. Soc. **67**, 617 (1945); Edsall, Foster, and Scheinberg, J. Am. Chem. Soc. **69**, 2731 (1947); J. T. Edsall and J. F. Foster, J. Am. Chem. Soc. **70**, 1860 (1948).
[w] V. N. Tsvetkov and A. Petrova, J. Tech. Phys. U.R.S.S. **12**, 423 (1942); V. N. Tsvetkov and E. Frisman, Acta. Physicochim. U.R.S.S. **20**, 61 (1945). The published details of the apparatus give only the gap widths and so it is impossible to calculate Ω_c and G_c. However, the formula used by these workers for calculating these values is identical with that used by Buchheim, Stuart and Menz and is incorrect. The anomalies mentioned may therefore be partly due to turbulence.

curve clearly shows that as s increases from zero to infinity so N_c changes from infinity, through a minimum value to a large value again, so that for a given radius of the outer cylinder a decrease in gap width may or may not allow an increase in the speed of revolution. Beyond a certain value of s, as s increases, so N_c increases but the advantage derived by having a large gap is counteracted by the fact that the gradient constancy factor increases, which means that the error in assuming the gradient to be constant across the gap is now larger.

Taylor[19] investigated the velocity variation across the gap and found (see Fig. 1b) that (a) the velocity distribution is similar to that obtained with stream line flow, (b) turbulence is not very effective at the outer layers and (c) due to the stabilizing effect of rotation there is a large velocity gradient across the gap. Consideration of Fig. 1b indicates that even when turbulence sets in a sharp break may not occur in the double refraction velocity gradient curves as the gradient does not differ greatly from that occurring with stream-line flow.

[19] G. I. Taylor, Proc. Roy. Soc. **A157**, 565 (1936).

III. APPLICATION TO EXISTING APPARATUS

(a) Inner Cylinder Rotating

The criterion for turbulence given by Eq. (1) has been applied to the apparatus of the majority of observers who have published the dimensions used. The results are summarized in Table I. Only in the apparatus of de Metz does the ratio s/R_1 exceed 0.33 and even then it is less than 0.71. The values of Ω_c, G_c, and N_c given,

can in all cases therefore, be taken as accurate to 1 percent. To find which of the results of the various authors are taken for conditions of laminar flow it is only necessary to calculate the value of ν for the liquid used. If then the number of revolutions per second of the cylinder exceeds that given in the seventh column, the results cannot be regarded as being of any quantitative significance.

(b) Outer Cylinder Rotating

The results in this case are given in Table II in which the values quoted for the critical constants are lower limits; in fact values given for $s/R_2>0.22$ are probably very much below the truth. It is immediately obvious that the values of G_c are very much greater than in the case of rotation of the inner cylinder although the constancy factor is greater. In only one case were the measurements made in the turbulent region and that is in the work of Winkler who studied benzene at gradients of 7710, 11,600 and 14,000 sec.$^{-1}$. If ν is assumed to be 0.0074 stokes[20] at 20°C the critical gradient is about 420 sec.$^{-1}$. Winkler was led to believe he was working in the laminar region for he refers only to Taylor's conclusion published in 1923[21] that flow should be laminar at all speeds of rotation. This conclusion was superseded by Taylor's 1936 paper,[22] to which Winkler made no reference.

IV. OPTICAL DESIGN

The concentric cylinders form only one component of the apparatus for investigating streaming double refraction. In designing an apparatus attention must also be paid to the optical components, for there are numerous optical sources of error, which, if ignored, are liable to introduce double refraction of the same order as that being measured. The main errors are (a) polarization at lenses, (b) polarization by reflection from metal surfaces (c) strain in the cover glasses through which the light leaves the liquid (d) the effect of heat developed in the liquid annulus.

An excellent survey of these effects has been given by Snellman[23] and they will, therefore, not be discussed in detail here. It is sufficient to say that no trouble occurs from (a) if no lenses are situated between the polarizing prisms; the effects of (b) are avoided if the optical system as developed by Bjornstahl[24,25] be used; the effects of (c) are avoided if the method of Frey-Wyssling and Weber[26] be used while the effect of the heat developed can be calculated by a theory developed by Bjornstahl.[27]

V. GENERAL REMARKS

From the results obtained in streaming double refraction experiments it is possible to get information on the properties of the particles or molecules in the liquid. One of the important parameters required to interpret the results is the viscosity under the prevailing conditions. The fact that the viscosity coefficient must be that measured under the prevailing conditions is especially important for non-Newtonian liquids or those exhibiting anomalous viscosity. Results derived from a combination

TABLE II. Dimensions of cylinders in apparatus used by various investigators and critical velocity gradients—outer cylinder rotating.

Author		Inner cylinder rad. (R_1) cm	Gap width (s) cm	Gradient constancy factor $\frac{s}{R_1}\frac{dG}{G}=-(10^{-3})$	Critical angular velocity Ω_c (ν-rad. sec.$^{-1}$)	Critical velocity gradient G_c (ν sec.$^{-1}$)	Critical No. of revs. N_c (ν r.p.s.)
Freundlich, Stapelfeldt and Zocher[a]	(a)	1.00	0.45	45.0	69000	222000	11000
	(b)	0.35	1.10	314.3	28000*	37000	4400
	(c)	0.55	0.90	163.6	34500*	55500	5500
von Muralt and Edsall[b]	(a)	0.91	1.09	119.8	21000*	38000	3300
	(b)	1.30	0.70	53.8	32000*	92000	5100
	(c)	1.80	0.20	11.1	17300	173000	2750
Robinson[c]		1.7275	1.135	65.7	13800*	35000	2200
Winkler[d]		5.06	0.2	3.9$_6$	2130	56000	340
Lawrence†,[e] *et al.*	(a)	0.20	0.192	96.0	600000*	1221000	95500
	(b)	0.32	0.072	22.5	707000	3849000	112500

* In these cases $s/R_2>0.22$: the critical values are probably very much higher than those given here.

[a] Freundlich, Stapelfeldt, and Zocher, Zeits. f. physik. Chemie **114**, 161, 190 (1925).

[b] A. L. von Muralt and J. T. Edsall, J. Biol. Chem. **89**, 351 (1930); Trans. Faraday Soc. **26**, 837 (1930).

[c] J. R. Robinson, Proc. Roy. Soc. **A170**, 519 (1939).

[d] E. Winkler, Zeits. f. Physik **118** (3-4), 232 (1941). See also E. Winkle and W. Kast, Naturwiss. **29**, 288 (1941).

† Some uncertainty exists as to the actual dimensions of this instrument. It would appear from the values of the gap width given by the authors that the figures given for the radii are actually those for the diameter. In any case there can be no question of turbulence occurring at the speeds used.

[e] Lawrence, Needham, and Shen, J. Gen. Physiol. **27**, 201 (1944).

[20] This value is one derived from measurements with a flow tube viscometer.

[21] G. I. Taylor, Phil. Trans. **A223**, 289 (1923).

[22] G. I. Taylor, Proc. Roy. Soc. **A157**, 546 (1936).

[23] O. Snellman, Ark. Kemi. Min. Geol. **19A** (30), 1 (1945).

[24] Y. Bjornstahl, J. Opt. Soc. Am. **29**, 201 (1939).

[25] Frey-Wyssling and Weber (Helv. Chim. Acta **24**, 278 (1949) suggested a vectorial method of correcting this effect but their method besides being laborious may not be valid owing to the assumptions used. If Bjornstahl's method is impracticable, the method suggested by J. T. Edsall and J. F. Foster (J. Am. Chem. Soc. **67**, 617 (1945)) in which the length of the light path between the cylinder and observer is increased so as to reduce the intensity of the reflected light, may be used.

[26] A. Frey-Wyssling and E. Weber, Helv. Chim. Acta **24**, 278 (1941).

[27] Y. Bjornstahl, Zeits. f. Physik **119**, 3-4, 245 (1942).

of double refraction measurements made in a concentric cylinder apparatus and viscosity measurements made in a flow-tube viscometer are subject therefore to errors, which may be large for anomalous liquids. This point seems to have been overlooked by the majority of observers and suggests the use of a streaming double refraction apparatus with rotating outer cylinder and the inner cylinder freely suspended so that the whole forms a Couette viscometer.

It might well be mentioned here that the usual equation employed for obtaining the viscosity coefficient in a Couette viscometer is one deduced on the assumption that the liquid in the annular gap is Newtonian. If this equation is used for a non-Newtonian liquid the viscosity coefficient so obtained is an "apparent" coefficient equal to that of a Newtonian liquid producing the same deflection under identical conditions in the particular apparatus used.[28]

From the point of view of obtaining high values of the velocity gradient without turbulence there is theoretically an advantage in employing the outer cylinder as rotor although this advantage becomes less marked as the gap width is narrowed. However, from the practical point of view the difficulty experienced in the optical system with a narrow gap is greater than with a large gap.

Finally it must be remembered that the criteria for turbulence as given by Taylor hold only for Newtonian liquids and only when turbulence is due solely to centrifugal forces. Also it would appear that the ratio of cylinder length to gap width should be large. Any external disturbance, such as vibration, will tend to change the conditions and turbulence may therefore occur before the critical values given by Eq. (1) or the curves of Fig. 3. This was mentioned to the author in a private communication by Sir Geoffrey Taylor. Some such disturbance has in fact been detected by Buchheim, Stuart, and Menz[29] for even when allowance has been made for their mistake in applying Taylor's criterion incorrectly, turbulence sets in prematurely. Furthermore solutions which have long rod-shaped particles (i.e. which exhibit rod double refraction) may show turbulence earlier than Taylor's criteria would indicate. This is in accordance with the fact that the anomalous viscosity of such solutions will tend to change the value of the critical velocity (or the critical value of Reynolds number) for which turbulence occurs. Such a case has been observed by Andrade and Lewis[30] with ammonium oleate, the critical value of the velocity gradient being found to be about 30 percent lower than that given by Taylor's criterion for homogeneous liquids. This is probably the explanation of the anomalies observed by Hatschek and Jane.[31–33]

These facts indicate that it would be advantageous to have some physical indication of the onset of turbulence rather than having to rely upon calculations. This is possible if one adopts as a test for the onset of turbulence the fact that the relationship between viscous torque on the stationary cylinder and the speed of rotation of the moving cylinder instead of being linear becomes irregular. A brief discussion of this criterion as applied by Couette[34] and Mallock[35] is given by Taylor.[36]

A gradual departure from the linear law must, however, be carefully considered before it is taken as being due to turbulence for if the liquid under investigation is anomalous the torque may increase less quickly than the speed of rotation.[37]

VI. CONCLUSIONS

Tables I and II provide means by which an estimation may be made of the significance of the results of published work.

Theoretical and practical advantages in using the outer cylinder as rotor instead of the inner one are the following: (a) A larger gap can be used without the risk of turbulence and consequently greater ease is experienced in the optical measurements. (b) With a larger gap the temperature rise in the liquid is less. As the viscosity varies appreciably with temperature, the temperature of the liquid should be measured and not assumed to be that of the cooling bath. For this purpose it is advisable to incorporate a thermocouple in the concentric cylinder apparatus which may conveniently be built in the surface of the stationary cylinder. (c) The instrument can be used as a Couette viscometer so that double refraction and viscosity measurements can be made simultaneously.[38] (d) The onset of turbulence may be directly detected by irregularities in the torque-speed curve.

The onset of turbulence in pure liquids may not be accompanied by an abrupt change in the double refraction across the gap, although some change should occur consistent with the observed velocity distribution existing across the gap. A similar result holds for anomalous liquids, although it is probable that the velocity distribution across the gap under turbulent conditions differs slightly from that for a pure liquid.

---

[28] E. Hatschek, *Viscosity of Liquids* (G. Bell and Sons Ltd., London, 1928), p. 191. A. C. Merrington, *Viscometry* (E. Arnold & Company, London, 1949), p. 106.

[29] Buchheim, Stuart, and Menz, Zeits. f. Physik **112**, (7–8), 407 (1939).

[30] E. N. da C. Andrade and J. W. Lewis, Kolloid. Zeits. **38**, 260 (1926).

[31] E. Hatschek and R. S. Jane, Kolloid. Zeits. **38**, 33 (1926).

[32] E. Hatschek, Kolloid. Zeits. **38**, 259 (1926).

[33] W. Ostwald and R. Auerbach, Kolloid. Zeits. **38**, 261 (1926).

[34] H. Couette, Ann. de. Chim. et de Phys. 6^{me} sér., 21 (1890).

[35] A. Mallock, Phil. Trans. **A187**, 41 (1896).

[36] G. I. Taylor, Phil. Trans. **A223**, 289 (1923).

[37] E. Hatschek, Kolloid. Zeits. **13**, 88 (1913). E. Hatschek and E. Humphrey, Proc. Phys. Soc. (London) **28**, 274 (1916). E. Hatschek and R. S. Jane, Kolloid. Zeits. **40**, 53 (1926). W. Ostwald, Kolloid. Zeits. **43**, 210 (1927).

[38] For a discussion of the end effects in rotational viscometers see C. H. Lindsley and E. K. Fischer [J. App. Phys. **18**, 988 (1947)]. The effect of the ends on the velocity distribution across the gap is treated by Taylor [Proc. Roy. Soc. **A157**, 572 (1936)].

In the light of the above discussion an instrument has been designed and is under construction; it is hoped that simultaneous measurements of viscosity and streaming double refraction will be possible.

VII. ACKNOWLEDGMENT

The author wishes to record the help and advice of Professor A. M. Taylor in whose laboratories research on streaming double refraction is being carried out.

Effective Stress and Effective Strain in Relation to Stress Theories of Plasticity*

G. N. WHITE, JR.** AND D. C. DRUCKER†
Brown University, Providence, Rhode Island

(Received April 3, 1950)

The correlation of experimental data and the proper description of the state and history of deformation of work-hardening materials has received much attention. An often desired objective is the plotting of a variable involving stress alone, a so-called *effective stress*, against a variable depending on the instantaneous strain or the history of strain, *effective strain*. Quantities such as maximum or octahedral shearing stress and strain or better an integrated effective strain increment have been employed. Also, simple effective stress definitions have been given for anisotropic as well as isotropic metals by Dorn, by Jackson, Smith, Lankford, and by previous investigators.

It is shown that the stress-strain relations of a recent theory of plasticity provide a very convenient means of defining and studying effective stress and effective strain. The strong theoretical and practical limitations of such concepts are indicated and the calculation of specific plastic work is discussed for both isotropic and anisotropic metals.

One of the major objectives of the paper is to stimulate critical experimental examination of as yet unproved assumptions in the theory of plasticity and in the correlation of experimental data.

INTRODUCTION

THE relation between stress and strain for a work-hardening material is given for any one simple loading program (tension, shear, or any loading with constant ratios of the stress components) by a single curve. When the results obtained with different loading programs are to be correlated, or when the state and history of deformation for complex loading paths is to be described[1,2] it is again desirable to obtain a single curve. What is most often sought is a plot of some function of the stresses, termed the effective stress, against a function of the strain or strain history, termed effective strain. The related problem of equivalent proof stress has also been treated.[3]

The results of combined loading experiments have often been correlated by intuitively chosen stress and strain variables. Curves showing maximum shear stress *vs.* maximum shear strain are sometimes used. However, plots of octahedral shearing stress against octahedral shearing strain usually produce somewhat better correlation and are most often employed at present.

More recent definitions of effective stress $\bar{\sigma}$ and effective strain $\bar{\epsilon}$ have been developed in an entirely different manner to correlate not only simple loading programs but also those in which the directions and ratios of the principal stresses are not constant. Use is made of a stress-strain relation and a definition of effective strain based on the increment of specific plastic work, i.e. plastic work per unit volume, $dW^p = \bar{\sigma} d\bar{\epsilon}$. This approach has been employed for both isotropic and anisotropic materials.[4–7] If the effective stress $\bar{\sigma}$ in $dW^p = \bar{\sigma} d\bar{\epsilon}$ is taken to be the octahedral shearing stress, the increment of effective strain $d\bar{\epsilon}$ is found to be a function of the strain increments alone. For other definitions of $\bar{\sigma}$ it is by no means clear that the effective strain increment so defined will bear any simple relation to the measured or computed strain increments nor, in fact, have any direct physical significance.

The objective of this paper is to discuss these questions and to show how incremental stress-strain relations derived in terms of a criterion of yield or of further plastic deformation (termed a loading function) can be used in a systematic way to develop consistent and reasonable definitions of effective stress and effective strain. In this paper the loading function is chosen to be a function of the stress components only as is the current practice in experimental research. The choice of such a loading function, characteristic of the so-called "stress" theories of plasticity, leads directly to the definition of effective stress in terms of the stress com-

* The conclusions presented in this paper were obtained in the course of research conducted under contract N7onr-358 sponsored jointly by the ONR and the Bureau of Ships.

** Research Associate in Applied Mathematics, Brown University.
† Now Professor of Engineering, Brown University.

[1] W. Lode, Forsch. Gebiete Ingenieurw. No. 303 (1928).
[2] E. A. Davis, J. App. Mech., Trans. ASME **65**, A187–196 (1943).
[3] J. Marin and R. L. Stanley, Am. Welding Soc. J., Supplement 745–805 (1940).
[4] C. Zener and J. H. Hollomon, Trans. A. S. M. **33**, 163–214 (1944).
[5] Jackson, Smith, and Lankford, Am. Inst. Mining Met. Eng., Tech. Pub. No. 240, Class E, Metals Tech. August (1948).
[6] J. E. Dorn, J. App. Phys. **20**, 15–20 (1949).
[7] J. C. Fisher, Trans. ASME. **71**, 349–356 (1949).

ponents and, after some manipulation, to effective strain in terms of the history of strain. It should be kept in mind, however, that experimental evidence shows that the loading function also depends explicity on the strain (see reference 16). Moreover, effective stress must, in general, include stress history or the path of loading rather than final values of stress only. Therefore, the concepts of effective stress and strain as now employed and discussed in the following pages can have only a limited, although very useful, range of validity.

SYMBOLS

Note that repeated letter subscripts denote summation and any dummy letter may be used. For example, $\sigma_{ii} \equiv \sigma_{kk} = \sigma_{11}+\sigma_{22}+\sigma_{33} \equiv \sigma_x+\sigma_y+\sigma_z$ and $\sigma_{ij}d\epsilon_{ij} \equiv \sigma_{mn}d\epsilon_{mn} = \sigma_{11}d\epsilon_{11}+\sigma_{12}d\epsilon_{12}+\cdots+\sigma_{31}d\epsilon_{31}+\sigma_{33}d\epsilon_{33} = \sigma_x d\epsilon_x+\frac{1}{2}\tau_{xy}d\gamma_{xy}+\cdots+\frac{1}{2}\tau_{xz}d\gamma_{xz}+\sigma_z d\epsilon_z$.

c, a constant chosen to make effective stress reduce to a standard stress in some special case.

f, the loading function or criterion of plastic deformation; here taken as a function of the stresses alone.

$g=g(f)$ or $h(f)$, the part of the proportionality factor G in the stress-strain relations which depends on f alone.

m, an arbitrary constant in the proportionality factor G.

n, Poisson's ratio.

r, degree of f in the stresses.

$s_{ij}=\sigma_{ij}-(\sigma_{kk}/3)\delta_{ij}$, the stress deviator tensor $s_x=\frac{1}{3}(2\sigma_x-\sigma_y-\sigma_z)$, $s_{xy}=\tau_{xy}$, etc.

$t_{ij}=s_{ik}s_{kj}-\frac{2}{3}J_2\delta_{ij}$, the deviation of the square of s_{ij}.

A_{ijkl}, coefficient of the stresses in an anisotropic quadratic loading function.

A_{ij}, constants in the general anisotropic quadratic loading function.

B_{ij}, the inverse of A_{ij}, i.e. $B_{ij}A_{jk}=\delta_{ik}$.

C_{ijkl}, anisotropic elastic constants in the generalized Hooke's law.

A, B, C, D, E, F, constants in an orthotropic loading function.

E, Young's modulus.

G, the factor of proportionality in stress-strain relations: here a function of stress.

$H=H[(K_2)^{\frac{1}{2}}, (|K_3|)^{\frac{1}{3}}]$, a function of K_2, K_3 used to define an increment of effective strain.

$I(f)=H/df$, a function of f alone used to define an increment of effective strain.

$J_1=\sigma_{kk}$, the first invariant of the stress tensor σ_{ij} equal to $\sigma_x+\sigma_y+\sigma_z$.

$J_2=\frac{1}{2}s_{ij}s_{ij}$, the second order invariant of s_{ij}, the stress deviator or $\frac{1}{6}[(\sigma_1-\sigma_2)^2+(\sigma_2-\sigma_3)^2+(\sigma_3-\sigma_1)^2]$ equal to $\frac{2}{3}\tau_0^2$.

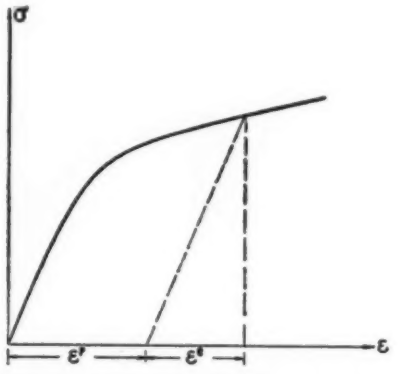


FIG. 1. Tension curve for a work hardening material.

$J_3=\frac{1}{3}s_{ij}s_{jk}s_{ki}$, the third order invariant of s_{ij}.

$K_2=d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p$, second invariant of $d\epsilon_{ij}{}^p$ equal to $(d\epsilon_x)^2+(d\epsilon_y)^2+(d\epsilon_z)^2+\frac{1}{2}(d\gamma_{xy})^2+\frac{1}{2}(d\gamma_{yz})^2+\frac{1}{2}(d\gamma_{zx})^2$.

$K_3=d\epsilon_{ij}{}^p d\epsilon_{jk}{}^p d\epsilon_{ki}{}^p$, third invariant of $d\epsilon_{ij}{}^p$.

$M_2=(\partial f/\partial\sigma_{ij})(\partial f/\partial\sigma_{ij})$, second invariant of $(\partial f/\partial\sigma_{ij})$, the gradient of f.

$M_3=(\partial f/\partial\sigma_{ij})(\partial f/\partial\sigma_{jk})(\partial f/\partial\sigma_{ki})$, third invariant of $(\partial f/\partial\sigma_{ij})$.

$R=(\frac{1}{2}M_2)^{\frac{1}{2}}/(\partial f/\partial\tau)$, the factor of difference between two stress-strain relations at a given value of f.

W^p, dW^p, specific plastic work and its increment.

δ_{ij}, the Kronecker delta equal to 0 for $i\neq j$, 1 for $i=j$.

$\bar{\epsilon}$, $\bar{\epsilon}^p$, effective strains, total and plastic.

$d\epsilon_{ij}=d\epsilon_{ij}{}^e+d\epsilon_{ij}{}^p$, increments of the total strain tensor ϵ_{ij} equal to the sum of the elastic and plastic increments with $d\epsilon_{11}=d\epsilon_x$, $d\epsilon_{12}=\frac{1}{2}d\gamma_{xy}$, $\cdots$, $d\epsilon_{33}=d\epsilon_z$.

ϵ_1, ϵ_2, ϵ_3, principal strains.

γ_0, octahedral shearing strain.

μ, ν, Lode's variables $\mu=(2\sigma_2-\sigma_1-\sigma_3/\sigma_1-\sigma_3)$, $\nu=(2d\epsilon_2-d\epsilon_1-d\epsilon_3/d\epsilon_1-d\epsilon_3)$.

$\bar{\sigma}$, effective stress.

σ_{ij}, stress tensor with components $\sigma_{11}=\sigma_x$, $\sigma_{12}=\tau_{xy}$, $\cdots$, $\sigma_{33}=\sigma_z$.

σ_1, σ_2, σ_3, principal stresses.

τ_0, octahedral shearing stress.

$\tau=\tau_{xy}$, shearing stress when $\tau_{yz}=\tau_{zx}=0$.

THE STRESS-STRAIN RELATION

The stress-strain relation to be used in this paper[8,9] is of the incremental or flow type which includes as special cases all explicit incremental laws now in use. A brief discussion of the development of the stress-strain relation will be given before the subject of effective stress and strain is treated. In this way the resulting definitions of effective stress and strain may be evaluated in the light of the assumptions made in deriving the stress-strain relation.

Figure 1 shows a typical tensile stress-strain curve for a work-hardening material. It is assumed that time and temperature effects are small enough to be disregarded. The stress-strain curve, which has positive slope, for the entire range considered represents then a series of states of stable static equilibrium. As no simple correlation between stress and strain exists for complex loading paths, the meaning of work-hardening for the general case requires consideration.

Work-hardening may be defined in the following manner: If for all additional sets of stresses which are slowly applied to and then removed from a block of stressed material in a given state of deformation by an external agency, the material remains in equilibrium and (a) positive work is done *by the external agency* during the application of the stresses and (b) the net work performed by it over the cycle of its application and removal is positive (or zero if no plastic deformation occurs), then the material is said to be work hardening. In other words, no work can be extracted in a cycle from the system including the block of material and the forces acting upon it.

The stress-strain relations which are derived in con-

[8] W. Prager, J. App. Phys. **20**, 235–241 (1949).

[9] D. C. Drucker, J. Coll. Sci., Rheology Issue 4, 299–311 (1949) D. C. Drucker, Quart. App. Math. **7**, 411–418 (Jan. 1950); D. C. Drucker, Trans. A. S. M. E. **71**, 587–592 (1949).

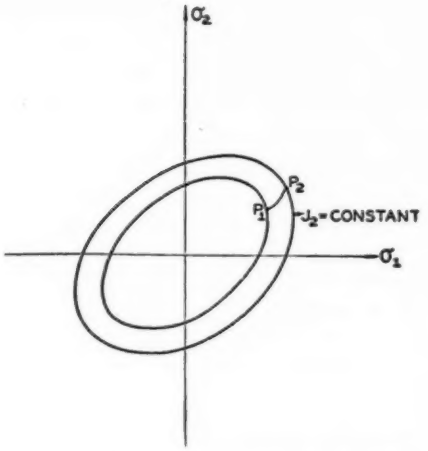


FIG. 2. A level curve of $f=J_2=\frac{1}{3}\tau_0^2$.

formity with the work hardening conditions are based on two assumptions:

(1) The relation between the increments of stress and strain is linear (all explicit stress-strain relations proposed so far adopt this assumption).

(2) A yield criterion or loading function f exists.

At any given state of plastic deformation a function of the stress components, $f(\sigma_{ij})=f(\sigma_x, \sigma_y, \cdots, \tau_{yz})$, exists such that further plastic deformation takes place only if f increases. In the general theory the value of the loading function f and also the form of f will depend on the plastic strain and plastic strain history.

In order to write the stress-strain relations compactly and to express basic ideas clearly, tensor notation will be employed. For example, the state of stress is represented by the single symbol σ_{ij} which stands for the stress tensor with the components $\sigma_{11}=\sigma_x$, $\sigma_{12}=\tau_{xy}$, $\cdots$, $\sigma_{33}=\sigma_z$. Moreover use of the tensor convention of summing over repeated subscripts, results in simplification of formulae and in ease of manipulation, e.g. $a_ib_i=a_1b_1+a_2b_2+a_3b_3$, and $a_{ii}=a_{11}+a_{22}+a_{33}$.

The work-hardening conditions and the two assumptions lead to an expression for the plastic strain increments $d\epsilon_{ij}^p$ in terms of the gradient of the loading function $\partial f/\partial\sigma_{ij}$, a proportionality factor G, and the stress increments $d\sigma_{kl}$. In tensor notation the relation is

$$d\epsilon_{ij}^p=G\frac{\partial f}{\partial\sigma_{ij}}\left(\frac{\partial f}{\partial\sigma_{kl}}d\sigma_{kl}\right). \tag{1}$$

The factor G is in general a function of the stress, strain, and strain history. The positive quantity $(\partial f/\partial\sigma_{kl})d\sigma_{kl}$ is the increase in the loading function resulting from stress changes only; for example $d\tau_0$, if τ_0, the octahedral shearing stress, is the criterion of loading. The total increment of strain $d\epsilon_{ij}$ is obtained by adding $d\epsilon_{ij}^e$, the elastic increment, to the plastic increment $d\epsilon_{ij}^p$. A general anisotropic form of Hooke's law, $d\epsilon_{ij}^e=C_{ijkl}d\sigma_{kl}$, or the usual isotropic one may be used.

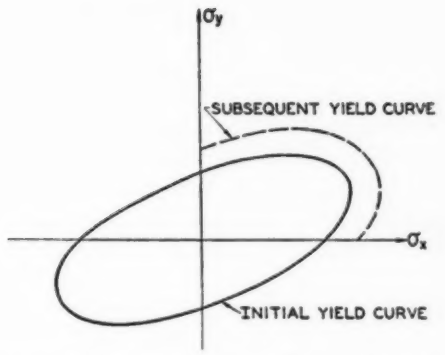


FIG. 3. Intrinsic anisotropy $f=f(\sigma)$, the flow stress in simple tension in the x-direction is not the same as for the y-direction.

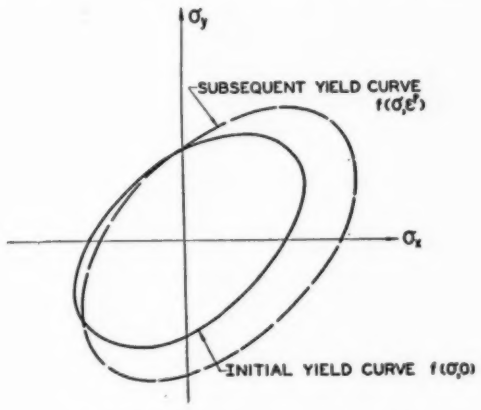


FIG. 4. Bauschinger effect—$f=f(\sigma, \epsilon^p)$.

The two resulting expressions are respectively

$$d\epsilon_{ij}=d\epsilon_{ij}^e+d\epsilon_{ij}^p=C_{ijkl}d\sigma_{kl}+G\frac{\partial f}{\partial\sigma_{ij}}\left(\frac{\partial f}{\partial\sigma_{kl}}d\sigma_{kl}\right) \tag{2}$$

and

$$d\epsilon_{ij}=\frac{(1+n)}{E}d\sigma_{ij}-\frac{n}{E}\delta_{ij}d\sigma_{kk}+G\frac{\partial f}{\partial\sigma_{ij}}\left(\frac{\partial f}{\partial\sigma_{kl}}d\sigma_{kl}\right), \tag{3}$$

where n is Poisson's ratio, E is Young's modulus, and δ_{ij} is the Kronecker delta, which is zero when $i\neq j$, unity when $i=j$. If the simplest loading function, the Mises' yield condition $f=J_2$ (shear strain energy or square of the octahedral shearing stress to within a multiplicative constant) is used, Eq. (3) in engineering notation is

$$\left.\begin{aligned} d\epsilon_x&=\frac{1}{E}[d\sigma_x-n(d\sigma_y+d\sigma_z)]\\ &\qquad+g(\tau_0)[\sigma_x-\tfrac{1}{2}(\sigma_y+\sigma_z)]d\tau_0\\ 2d\epsilon_{xy}=d\gamma_{xy}&=\frac{2(1+n)}{E}d\tau_{xy}+3g(\tau_0)\tau_{xy}d\tau_0 \end{aligned}\right\}. \tag{4}$$

These forms have been used by Nadai[10] and later by Dorn.[11]

Figure 2, a plot of Mises' loading function J_2 = constant for a two-dimensional stress system, reveals many of the features of the more general loading functions. According to the postulates for the stress-strain relations, if the stresses change so as to move the point P_1 representing them outside of the ellipse J_2 = constant on which the point is located, plastic deformation will take place and a new ellipse is established through the new point P_2. If the stresses change so as to move the point inside or on the ellipse, only elastic deformation can occur. In general, at any stage of plastic deformation, the loading function $f(\sigma_{ij})$ may be regarded as a surface in a multi-dimensional space with coordinates represented by all those stress components σ_{ij} which are not kept zero during the test. For a given state of stress at a point in the material, it is a surface which bounds all those states of stress which can be reached without additional plastic deformation taking place. As the increment of plastic work is necessarily positive, the surface f must be convex with respect to the origin as illustrated by Fig. 2.

An advantage of the general stress-strain relations Eqs. (2) and (3) is that anisotropic materials may be treated almost as easily as isotropic. Two types of anisotropy are (a) an initial anisotropy which is not altered appreciably by subsequent deformation, the loading function depending explicitly upon the stress components alone $f=f(\sigma_{ij})$, and (b) anisotropy produced by the plastic deformation being considered, the loading function depending explicitly on the components of plastic strain as well as stress, e.g. $f=f(\sigma_{ij}, \epsilon_{ij}{}^p)$. The first type of anisotropic loading function may be useful for a material which has been strongly cold worked prior to testing; the second for a material which is initially isotropic and exhibits a marked Bauschinger effect. Two-dimensional illustrations for particular examples of the two types of anisotropy are given in Figs. 3 and 4.

As f is here taken to depend explicitly on stress alone, the most general form of isotropy considered is that produced by using a loading function which is a function of the stress invariants J_1, J_2, and J_3. The octahedral shearing stress, $\tau_0=(\frac{2}{3}J_2)^{\frac{1}{2}}$, is a special case where $J_1=\sigma_{kk}$, the sum of the principal stresses, and J_3, which is a weighting factor for shearing stresses, are assumed to have no influence. The stress-strain relation (3) for $f=J_2=\frac{3}{2}\tau_0{}^2$ and $G=G(f)$ has been given in engineering notation, Eq. (4). In the simple tensor form, the plastic strain increment is

$$d\epsilon_{ij}{}^p=G(J_2)s_{ij}dJ_2=h(\tau_0)(\sigma_{ij}-(\sigma_{kk}/3)\delta_{ij})d\tau_0 \quad (5)$$

as s_{ij}, the stress deviator, is $\sigma_{ij}-(\sigma_{kk}/3)\delta_{ij}$. G and h are functions of J_2, or equivalently, of τ_0. A material obeying the relation (5) which uses J_2 as the loading function is called isotropic work hardening as it exhibits no directional properties when unloaded. There is no way of distinguishing the type of loading which produced the work hardening. All tensile specimens cut from the strained material will have the same yield point no matter what their orientation. Also, if the material is loaded a given amount in tension in any direction the extent of yielding or the value of f will be the same as for any other direction. Although the relation (5) is isotropic in the sense described, writing all the terms shows that when the material is under load (5) has the same form as a general anisotropic Hooke's law:

$$d\epsilon_x=G_1d\sigma_x+G_2d\sigma_y+G_3d\sigma_z+G_4d\tau_{xy}+G_5d\tau_{yz}+G_6d\tau_{zx}, \quad (6)$$

where the G's depend on stress alone. However, the anisotropy is produced entirely by the existing state of stress; it does not subsist after unloading.

Experimental evidence clearly indicates the need for more general stress-strain relations. The next extension is to retain the assumption of isotropic work hardening but include the third invariant of the stresses J_3 as well as J_2. The form of the relation (1) becomes then

$$d\epsilon_{ij}{}^p=G\left\{\frac{\partial f}{\partial J_2}s_{ij}+\frac{\partial f}{\partial J_3}t_{ij}\right\}df, \quad (7)$$

where t_{ij} is the deviation of the square of s_{ij}, $s_{ik}s_{kj}-\frac{2}{3}J_2\delta_{ij}$, and G is a function of stress.

If an intrinsically anisotropic stress-strain relation is needed, the loading function can no longer be a function of the stress invariants J_2 and J_3 but must depend directly on stress components referred to fixed axes. A simple convenient choice is the anisotropic quadratic form with constant coefficients $f=A_{ijkl}\sigma_{ij}\sigma_{kl}$ which reduces to J_2 in the isotropic case if the sum of the principal stresses is without influence. The stress-strain relation then obtained from Eq. (1) is

$$d\epsilon_{ij}{}^p=2GA_{ijkl}\sigma_{kl}df \quad (8)$$

where G is a function of stress. In expanded form using $\tau_{ij}=\tau_{ji}$

$$\left.\begin{aligned} d\epsilon_x{}^p&=2Gdf\{A_{1111}\sigma_x+2A_{1112}\tau_{xy}+A_{1122}\sigma_y \\ &\quad +2A_{1123}\tau_{yz}+A_{1133}\sigma_z+2A_{1131}\tau_{zx}\} \\ 2d\epsilon_{xy}{}^p=d\gamma_{xy}{}^p&=4Gdf\{A_{1211}\sigma_x+2A_{1212}\tau_{xy}+A_{1222}\sigma_y \\ &\quad +2A_{1223}\tau_{yz}+A_{1233}\sigma_z+2A_{1213}\tau_{zx}\} \end{aligned}\right\} \quad (9)$$

where the A's are just numbers. For a material in which there exist planes or axes with respect to which the anisotropy is symmetric, the general quadratic loading function and the corresponding stress-strain relation may be simplified. If for example the material is orthotropic and the principal axes of stress coincide with the lines of intersection of the three mutually orthogonal planes of symmetry taken as coordinate planes, the loading function is $f=a_{ij}\sigma_i\sigma_j$ $(i,j=1, 2, 3)$ where the stresses σ_i are principal stresses $\sigma_1=\sigma_x$ etc.

[10] A. Nadai, J. App. Phys. 8, 205–212 (1937).
[11] Cunningham, Thompson, and Dorn, Am. Soc. Test. Mat. 47, 546–553 (1947).

A typical plastic strain increment is then

$$d\epsilon_1^p = 2Gdf(a_{11}\sigma_x + a_{12}\sigma_y + a_{13}\sigma_z). \quad (10)$$

Eqs. (9) and (10) are stress-strain relations used by Dorn, by Jackson, Smith, and Lankford, and by Fisher.

More elaborate forms of f may easily be written as required by experimental data.

EFFECTIVE STRESS AND EFFECTIVE STRAIN

Effective stress $\bar{\sigma}$ is by definition a quantity which determines whether or not further plastic deformation takes place. Since it must be defined consistently with the stress σ in a simple tension test of a work hardening material, Fig. 1, $\bar{\sigma}$ must be a positive increasing function of the stresses during plastic deformation. For stress theories of plasticity, $f=f(\sigma_{ij})$, the loading function f also by definition determines whether additional plastic deformation will occur. It too is a positive increasing function as long as plastic deformation continues, and if unloading occurs, plastic deformation is not resumed until the highest previous value of f is exceeded. For stress theories there is then no real choice in the definition of the effective stress $\bar{\sigma}$. The loading function f is the effective stress to some power and to within an arbitrary constant; for example $\tau_0=(\frac{2}{3}J_2)^{\frac{1}{2}}$ or $\bar{\sigma}=(3J_2)^{\frac{1}{2}}$.[10,11]

On the other hand the definition of an effective strain $\bar{\epsilon}$ related to the effective stress is not made so easily. The method now used is to arbitrarily define the permanent effective strain $\bar{\epsilon}^p$ by

$$dW^p = \bar{\sigma}d\bar{\epsilon}^p, \quad (11)$$

where dW^p is the increment of specific plastic work of deformation.[4-7] This definition is based on the belief that the plastic work is a fundamental measure of the amount of plastic deformation.

The definition (11) of $\bar{\epsilon}$ in terms of plastic rather than total work is made to avoid considering increases of total work in which plastic deformation plays no part. Such a situation arises when the loading path is one for which $f=$constant. In this case the plastic increments of strain (1) are zero whereas the elastic strain increments are not.

A second basic method is the use of the stress-strain relations, Eq. (1), to find an integral of the strain increments which is exactly or quite closely a function of the loading function f or effective stress $\bar{\sigma}$ only. The integral of the strain increments is then defined to be the effective strain.

The second method is an attempt to define an effective strain which has a direct physical meaning in terms of measured strain and strain history. The first method does not in general lead to an effective strain which is related to strain as ordinarily defined. The effective strain increment cannot be computed from the strains and their history only, as may be seen when Eq. (11) is expressed fully in terms of the stresses and strain increments by means of Eq. (1).

$$dW^p = \bar{\sigma}d\bar{\epsilon}^p = \sigma_{ij}d\epsilon_{ij}^p = \sigma_{ij}G\frac{\partial f}{\partial \sigma_{ij}}\left(\frac{\partial f}{\partial \sigma_{kl}}d\sigma_{kl}\right). \quad (12)$$

It should be kept in mind that the concept of effective stress $\bar{\sigma}$ and consequently of effective strain $\bar{\epsilon}^p$ as presented here is only valid for materials with negligible Bauschinger and similar effects, or the more usual case of paths of loading which do not reveal such effects directly. Figure 4 shows that to account for Bauschinger effect the loading function must depend on the plastic strains as well as on stress. An effective stress which depends on stress only is then no longer possible as consideration of alternating tension and compression clearly shows.

In the following sections the expressions for $\bar{\sigma}$ and $\bar{\epsilon}^p$ will be given explicit forms in accordance with certain other features of the loading functions and stress-strain relations dictated by experimental results.

ISOTROPIC THEORIES

The simplest form of stress-strain relation for a material which work hardens isotropically, Eq. (5), will be used first as an example of both methods for defining the effective strain $\bar{\epsilon}^p$ in terms of the effective stress $\bar{\sigma}$. In this case $\bar{\sigma}$ may be defined as $(f)^{\frac{1}{2}}$ or $(J_2)^{\frac{1}{2}}$ multiplied by any constant. Choosing $\bar{\sigma}=(3J_2)^{\frac{1}{2}}$ makes $\bar{\sigma}=\sigma_x$ for uniaxial tension in the x-direction. The resulting $\bar{\epsilon}^p$ is then the same as Dorn's.[11]

Using the first method and noting that $(\partial J_2/\partial\sigma_{ij})=s_{ij}$ and $(d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}}=G(2J_2)^{\frac{1}{2}}dJ_2$ gives

$$\begin{aligned}\bar{\sigma}d\bar{\epsilon}^p = dW^p &= \sigma_{ij}d\epsilon_{ij}^p = \sigma_{ij}G(J_2)(\partial J_2/\partial\sigma_{ij})dJ_2 \\ &= (3J_2)^{\frac{1}{2}}(\tfrac{2}{3}d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}} = \bar{\sigma}(\tfrac{2}{3}d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}}. \quad (13)\end{aligned}$$

Consequently the effective strain increment is

$$\begin{aligned}d\bar{\epsilon}^p = (\tfrac{2}{3}d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}} \equiv (\tfrac{2}{3})^{\frac{1}{2}}\big[&(d\epsilon_x^p)^2+(d\epsilon_y^p)^2 \\ &+(d\epsilon_z^p)^2+\tfrac{1}{2}(d\gamma_{xy}^p)^2+\tfrac{1}{2}(d\gamma_{yz}^p)^2+\tfrac{1}{2}(d\gamma_{zx}^p)^2\big]^{\frac{1}{2}}, \quad (14)\end{aligned}$$

as $d\gamma_{xy}=2d\epsilon_{xy}$, etc. and $d\epsilon_{xy}=d\epsilon_{yx}$.

On the other hand, the second method uses the stress-strain relation Eq. (5) directly to form an increment of effective strain. The simplest combination of the plastic strain increments which is positive increasing and has the correct dimension for an effective strain increment is $(d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}}$ so that with $(\frac{2}{3})^{\frac{1}{2}}$ inserted arbitrarily to make $\bar{\epsilon}^p=\epsilon_z^p$ for uniaxial tension in the x-direction

$$d\bar{\epsilon}^p = (\tfrac{2}{3})^{\frac{1}{2}}(d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}} = (\tfrac{2}{3})^{\frac{1}{2}}G(J_2)(2J_2)^{\frac{1}{2}}dJ_2 = G^*(\bar{\sigma})d\bar{\sigma} \quad (15)$$

which shows that the integrated effective strain $\bar{\epsilon}^p$ is a function of the effective stress $\bar{\sigma}$. Once the factor $G(J_2)$ has been found for a given material by matching the stress-strain relation, Eq. (5), to an experimental curve, the formula for $\bar{\epsilon}^p$ in terms of $\bar{\sigma}$ is completely determined. In this approach also the constant $\sqrt{3}$ in $\bar{\sigma}$ may be inserted if desired to make $\bar{\sigma}=\sigma_x$ for simple tension in the x-direction.

Before the definition of $\bar{\epsilon}^p$ is extended by means of each of the two methods to materials requiring more complicated loading functions, two features which account for the simplicity of the preceding result should be noted. The first is that the loading function $f=J_2=\frac{3}{2}\tau_0^2$ is homogeneous (quadratic) in the stresses; the second, that G is taken to depend on the loading function only. If these two properties are assumed to hold for other stress loading functions of degree r in the stresses and for the corresponding stress-strain relations, the expression for the specific plastic work is found to be a function of f

$$dW^p=\sigma_{ij}G(f)\frac{\partial f}{\partial\sigma_{ij}}df=rfG(f)df, \tag{16}$$

which, if true, would mean that when plastic deformation takes place the plastic work is independent of the path of loading.

Some evidence that f is a homogeneous function of the stresses exists in the results of Fraenkel's tube tests.[12] It was found that for a tube loaded in tension and internal pressure applied in constant ratio, $\mu-\nu$, the difference between Lode's variables, is quite constant for octahedral shear strains larger than 0.02 percent (see discussion[13]). Since ν is given by

$$\nu=\frac{2d\epsilon_2^p-d\epsilon_1^p-d\epsilon_3^p}{d\epsilon_1^p-d\epsilon_3^p}=\frac{2(\partial f/\partial\sigma)_2-(\partial f/\partial\sigma)_1-(\partial f/\partial\sigma)_3}{(\partial f/\partial\sigma)_1-(\partial f/\partial\sigma)_3}, \tag{17}$$

homogeneity means constancy of ν as found in the experiment.

A lack of dependence of the dissipated work on the path of loading is a rather surprising result for a nonconservative system. However, Schmidt[14] has shown that, for tubes tested under tension and torsion applied in ratio, equally satisfactory correlation is obtained by plotting $f=\tau_0$ *vs.* work and $f=\tau_0$ *vs.* γ_0. On the other hand, a study of the results of Taylor and Quinney[15] indicates that, in their tests at least, the work dissipated is strongly path dependent and does not depend on the loading function alone. This is one of the most important questions still unanswered and much experimental work remains to be done.

When the effective stress $\bar{\sigma}$ and strain $\bar{\epsilon}^p$ of the simple J_2 or τ_0 isotropic theory do not give satisfactory correlation of experimental data,[16] definitions of $\bar{\sigma}$ and $d\bar{\epsilon}^p$ which depend on a stress-strain relation including the third stress invariant J_3, Eq. (7), should be tried. The necessity for including J_3 to fit experimental results is apparent from the fact that the J_2 theory gives straight μ *vs.* ν curves, i.e. $\mu=\nu$, whereas experiments show ν to be numerically less than μ except at special points. For isotropic loading functions this means that f depends on both J_2 and J_3. A possible inference is that the effective strain include not only the second strain increment invariant $K_2=d\epsilon_{ij}^p d\epsilon_{ij}^p$, as in the J_2 theory, but also the third invariant $K_3=d\epsilon_{ij}^p d\epsilon_{jk}^p d\epsilon_{ki}^p$. However, it will be shown that $(K_2)^{\frac{1}{2}}$ alone provides a satisfactory effective strain increment for many correlations of experimental data.

In the more general isotropic theory $f=f(J_2, J_3)$ where f is homogeneous of degree r in the stresses. Then $\bar{\sigma}=\bar{\sigma}(J_2, J_3)=cf^{1/r}$ where c is a constant which may be adjusted arbitrarily, as before, to make $\bar{\sigma}=\sigma_x$ in the case of simple tension. The first definition of the effective strain $\bar{\epsilon}^p$ by means of $dW^p=\bar{\sigma}d\bar{\epsilon}^p$ is useful but one without direct physical meaning as it is a function of stress as well as strain and cannot be computed from the measurement of strain and strain history along. This approach is equivalent to plotting plastic work against effective stress.[14] On the other hand, the second method for defining the effective strain using the stress-strain relations does not necessarily load to an $\bar{\epsilon}^p$ which is precisely dependent on the loading function f alone. Hence one may have to choose an $\bar{\epsilon}^p$ which is very closely related to, but is not truly a function of, $\bar{\sigma}$ alone. It will, in fact, be shown that for $f(J_2, J_3)$ the effective strain as defined for the simpler J_2 theory is satisfactory for experimental purposes providing, of course, that a J_2, J_3 theory is adequate for the data.

Consider the definition of $d\bar{\epsilon}^p$ by

$$\left(\frac{3}{2}\right)^{\frac{1}{2}}d\bar{\epsilon}^p=(d\epsilon_{ij}^p d\epsilon_{ij}^p)^{\frac{1}{2}}\equiv(K_2)^{\frac{1}{2}}=G(f)\left(\frac{\partial f}{\partial\sigma_{ij}}\frac{\partial f}{\partial\sigma_{ij}}\right)^{\frac{1}{2}}df=G(f)(M_2)^{\frac{1}{2}}df \tag{18}$$

where M_2 is the second invariant of $\partial f/\partial\sigma_{ij}$, the gradient of f. For the isotropic loading functions under consideration Eq. (18) may be written

$$\left(\frac{3}{2}\right)^{\frac{1}{2}}d\bar{\epsilon}^p=G(f)\left[2J_2\left(\frac{\partial f}{\partial J_2}\right)^2+6J_3\frac{\partial f}{\partial J_2}\frac{\partial f}{\partial J_3}+\frac{2}{3}J_2^2\left(\frac{\partial f}{\partial J_3}\right)^2\right]df. \tag{19}$$

It is apparent from (19) that in general this choice of $d\bar{\epsilon}^p$ is not a function of f and df alone and cannot be integrated to give $\bar{\epsilon}^p$ as a function of f alone. Since this isotropic theory includes the effect of J_3, a definition of

[12] S. J. Fraenkel, J. App. Mech., Trans. ASME **70**, 193–200 (1948).
[13] D. C. Drucker and W. Prager, discussion of reference 12, J. App. Mech., Trans. ASME **71**, 101–102 (1949).
[14] R. Schmidt, Ingenieur Archiv. **3**, 215–235 (1932).
[15] G. I. Taylor and H. Quinney, Phil. Trans. Roy. Soc. **A230**, 323–362 (1931).
[16] D. C. Drucker, J. App. Mech., Trans. ASME **71**, 349–357 (1949).

$d\bar{\epsilon}^p$ may be tried which includes the third invariant K_3 as well as K_2, e.g.,

$$d\bar{\epsilon}^p = H([K_2]^{\frac{1}{2}}, [|K_3|]^{\frac{1}{3}}). \tag{20}$$

While neither K_2 nor K_3 are functions of f and df alone, the function $H([K_2]^{\frac{1}{2}}, [|K_3|]^{\frac{1}{3}})$ may be equivalent to a function I of f times df

$$d\bar{\epsilon}^p = H([K_2]^{\frac{1}{2}}, [|K_3|]^{\frac{1}{3}}) = I(f)df \tag{21}$$

and therefore may be integrated to give $\bar{\epsilon}^p = \bar{\epsilon}^p(f) = \bar{\epsilon}^p(\bar{\sigma})$. Although a relation of this type between H and I must always exist, $I(f)$ is given by an infinite series in J_2 and J_3, except in degenerate cases such as $f = J_2$. Definitions of $d\bar{\epsilon}^p$ of the type of Eq. (21) are therefore of little practical value. A restricted extension of the second method for defining $\bar{\epsilon}^p$ for $f = f(J_2, J_3)$ and $G = G(f)$

$$\bar{\epsilon}^p = \bar{\epsilon}^p\left(\int (K_2)^{\frac{1}{2}}dt, \quad \int (|K_3|)^{\frac{1}{3}}dt\right) = \bar{\epsilon}^p(f) \tag{22}$$

has the same defects as the definition of Eq. (21); a relation between the two integrals and f in (22) does exist but cannot generally be found in a simple, useful form.

These general procedures fail to give a useful $\bar{\epsilon}^p$, but a satisfactory $\bar{\epsilon}^p$ can be developed by means of the stress-strain relations Eq. (1) as the following example shows. Suppose that in Eq. (1) the function G is modified slightly to

$$G = g(f)/(M_2)^{\frac{1}{2}} \tag{23}$$

instead of $G = G(f)$ (as yet no conclusive experimental evidence exists for or against such a choice). The plastic strain increments are then

$$d\epsilon_{ij}{}^p = \frac{g(f)}{(M_2)^{\frac{1}{2}}} \frac{\partial f}{\partial \sigma_{ij}} df. \tag{24}$$

Squaring each term and adding gives

$$d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p = g^2(f)(df)^2, \tag{25}$$

so that $(d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p)^{\frac{1}{2}} = g(f)df$ when integrated will give a function of f alone. Thus any constant times $(d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p)^{\frac{1}{2}}$ is an effective strain increment; its integral $\bar{\epsilon}^p$ depends on f or the effective stress $\bar{\sigma}$ alone. The function G is easily modified to produce more complicated definitions of the effective strain in which both K_2 and K_3 appear. If G is

$$G = h(f)/(M_2)^{\frac{1}{2}} + m(|M_3|)^{\frac{1}{3}} \tag{26}$$

where

$$M_2 = \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{ij}}, \quad M_3 = \frac{\partial f}{\partial \sigma_{ij}} \frac{\partial f}{\partial \sigma_{jk}} \frac{\partial f}{\partial \sigma_{ki}},$$

and m is a constant, then the effective strain may be taken as any constant times

$$d\bar{\epsilon}^p = h(f)df = (K_2)^{\frac{1}{2}} + m(|K_3|)^{\frac{1}{3}}, \tag{27}$$

which again may be integrated directly. Still more general combinations, of course, may be derived.

These altered forms of the stress-strain relation are basically different from the previous ones for which $G = G(f)$. The work dissipated W^p is no longer dependent on the final state of stress or f alone. At present there is no definite indication as to which type of assumption is closer to physical reality. Experiment alone must decide, in each individual case, which stress-strain relation is best. As mentioned previously, the process of plastic deformation is irreversible so that even limited path independence $W^p = W^p(f)$ would be quite startling.

Fortunately, as the next section demonstrates, comparison of the results obtained with the various forms of G shows that the difference between $\bar{\epsilon}^p$ *vs.* $\bar{\sigma}$ curves in which $d\bar{\epsilon}^p$ is defined by means of the different factors is small. *Therefore $d\bar{\epsilon}^p$ can be chosen as $(\frac{2}{3})^{\frac{1}{2}}(d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p)^{\frac{1}{2}}$ for materials which work harden isotropically in the range considered.*

THE DEPENDENCE OF THE EFFECTIVE STRAIN $\bar{\epsilon}^p$ ON THE FACTOR G

It having been demonstrated that the effective strain $\bar{\epsilon}^p = \int (\frac{2}{3})^{\frac{1}{2}}(d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p)^{\frac{1}{2}}$ is a function of f or $\bar{\sigma}$ when the factor G in the stress-strain relation is given by $g(f)/(M_2)^{\frac{1}{2}}$ whereas $d\bar{\epsilon}^p$ is not in general a function of f along when $G = G(f)$, it is desirable to find a way of estimating the difference in $\bar{\epsilon}^p$ according to the two different types of G.

The two stress-strain relations may be written

$$d\epsilon_{ij}{}^{p(A)} = G(f)\frac{\partial f}{\partial \sigma_{ij}} df \tag{28}$$

$$d\epsilon_{ij}{}^{p(B)} = \frac{g(f)}{(M_2)^{\frac{1}{2}}} \frac{\partial f}{\partial \sigma_{ij}} df. \tag{29}$$

If the two relations are each matched to a given experimental shear stress *vs.* shear strain curve, τ *vs.* γ, it is readily found that, for an increment of f at any value of f, the increments of plastic strain for any loading path are related to each other quite simply by

$$d\epsilon_{ij}{}^{p(A)} = \frac{(\frac{1}{2}M_2)^{\frac{1}{2}}}{(\partial f/\partial \tau)} d\epsilon_{ij}{}^{p(B)} = R d\epsilon_{ij}{}^{p(B)} \tag{30}$$

where $R = (\frac{1}{2}M_2)^{\frac{1}{2}}/(\partial f/\partial \tau)$. The maximum difference between the two strains can be computed for a given loading function and given types of loading by finding the deviation of R from unity.

For the extreme case of the particular loading function $f = J_2{}^3 - (9/4)J_3{}^2$ in which J_3, the shearing stress weighting factor, is much more prominent than ordinarily demanded by experimental data, the range of variation of the factor of difference R is only $0.935 \leq R \leq 1$. Figure 5 shows the maximum difference in the

tension curves for the two laws after they have been matched in shear. The curve A is obtained by multiplying the strain value for the B curve by $R=0.935$. If instead of a tension curve, the effective stress-strain curve $\bar{\sigma}$ *vs.* $\bar{\epsilon}^p$ is desired, the same method is applied since $d\bar{\epsilon}^{p(A)}=(\frac{2}{3})^{\frac{1}{2}}(d\epsilon_{ij}{}^{p(A)}d\epsilon_{ij}{}^{p(A)})^{\frac{1}{2}}=Rd\bar{\epsilon}^{p(B)}$ according to the second method of defining effective strain.

The difference between the plots of $\bar{\sigma}$ *vs.* $\bar{\epsilon}^p$ for any other loading path, no matter how complicated, is less than for simple tension. Therefore, if effective strain $\bar{\epsilon}^p$ is to be plotted as a function of $\bar{\sigma}$, one may use $\bar{\epsilon}^p=(\frac{2}{3})^{\frac{1}{2}}\int(d\epsilon_{ij}{}^p d\epsilon_{ij}{}^p)^{\frac{1}{2}}$ with the assurance that even for extreme cases a small variation only occurs between a plot for which Eq. (28) applies and one for which Eq. (29) is correct. The effective strain, so defined, may be used to calculate the plastic work done by $W^p=\int\bar{\sigma}d\bar{\epsilon}^p$ within close limits providing, of course, that the assumption of isotropic work hardening is sufficiently good for the path of loading studied.

It is interesting to note that in the customary tests on thin walled tubes in which all stresses are increased in ratio, the simple $\bar{\epsilon}^p$ defined above is a constant times the octahedral shearing strain. If the material of the tube is reasonably isotropic at the start, all such tests, no matter what the ratios chosen, will give very closely

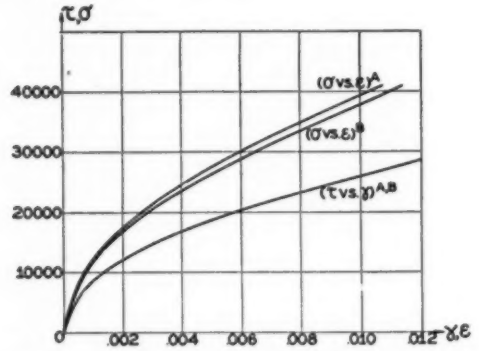


FIG. 5. Predicted σ *vs.* ϵ curves for A and B types of stress-strain relations based on a τ *vs.* γ curve of a severe test with the effect of J_3 overly prominent.

a single curve when the coordinates are chosen as $\bar{\sigma}(J_2, J_3)$ and $\bar{\epsilon}^p$ as the octahedral shearing strain.[16]

The expressions for effective stress $\bar{\sigma}=\bar{\sigma}(f)$ and effective strain $\bar{\epsilon}^p$ developed so far have been for isotropic work hardening and will not correlate experimental information from initially anisotropic materials. The general definitions of $\bar{\sigma}$ and $\bar{\epsilon}^p$, of course, also apply in the anisotropic case, but there the proper anisotropic stress-strain relations must be used.

ANISOTROPIC THEORY

As stated, the "stress" theories of plasticity $f=f(\sigma_{ij})$ with which this paper is concerned can only treat anisotropic materials which are free from Bauschinger effects for the loading being considered. The anisotropy is regarded as intrinsic in the material and is manifested in the loading function by asymmetry.

The work of Dorn,[6] of Fisher,[7] of Jackson, Smith and Lankford,[5] and of Hill[17] indicates that for many anisotropic materials, a loading function quadratic in the stresses is almost as satisfactory as J_2 is for the isotropic case. For such a loading function the stress-strain relation is given by Eq. (8)

$$d\epsilon_{ij}{}^p=2GA_{ijkl}\sigma_{kl}df. \tag{8}$$

As an example, the case of orthotropic symmetry in a material for which hydrostatic pressure does not influence plastic deformation and G is a function of f will be considered. The loading function f can be written as

$$\begin{aligned}2f=&A(\sigma_x-\sigma_y)^2+B(\sigma_y-\sigma_z)^2+C(\sigma_z-\sigma_x)^2+3D\tau_{xy}{}^2\\&+3D\tau_{yz}{}^2+3E\tau_{yz}{}^2+3E\tau_{zy}{}^2+3F\tau_{zx}{}^2+3F\tau_{zx}{}^2,\end{aligned} \tag{31}$$

where A, B, C, D, E, F are coefficients of anisotropy which become unity in the isotropic case. As f multiplied by any constant is just as satisfactory a loading function, one of the coefficients is not needed and can be taken as unity. The form (31) is employed purely for the sake of symmetry not generality.

The stress-plastic strain relations $d\epsilon_{ij}{}^p=G(\partial f/\partial\sigma_{ij})df$ become

$$\begin{aligned}d\epsilon_x{}^p&=G(\partial f/\partial\sigma_x)df=Gdf[A(\sigma_x-\sigma_y)-C(\sigma_z-\sigma_x)]\\d\epsilon_y{}^p&=G(\partial f/\partial\sigma_y)df=Gdf[B(\sigma_y-\sigma_z)-A(\sigma_x-\sigma_y)]\\d\epsilon_z{}^p&=G(\partial f/\partial\sigma_z)df=Gdf[C(\sigma_z-\sigma_x)-B(\sigma_y-\sigma_z)]\\d\epsilon_{xy}{}^p&=1/2d\gamma_{xy}{}^p=G(\partial f/\partial\tau_{xy})df=Gdf3D\tau_{xy}\end{aligned} \tag{32}$$

etc.

The ratios of the constants may be determined by simple tests. For example, under uniaxial tension, σ_x,

$$d\epsilon_y{}^p/d\epsilon_x{}^p=A/C, \quad \text{or} \quad d\epsilon_y{}^p/d\epsilon_x{}^p=-A/(A+C). \tag{33}$$

As f is quadratic the effective stress $\bar{\sigma}$ must be taken as $c(f)^{\frac{1}{2}}$ just as for the isotropic J_2. However, the choice of c which makes $\bar{\sigma}=\sigma_x$ for uniaxial tension in the x-direction does not make $\bar{\sigma}=\sigma_y$ for uniaxial σ_y.

A definition of effective strain corresponding to the anisotropic loading function (31) is obtained quite easily by either method considered. Here, in this special form, as in the J_2 isotropic quadratic theory, $dW^p=\bar{\sigma}d\bar{\epsilon}^p$ defines a $d\bar{\epsilon}^p$ in terms of the plastic strain increments alone. This is true because $dW^p=\sigma_{ij}d\epsilon_{ij}{}^p=G2fdf$ $=(G(2/c)(f)^{\frac{1}{2}}df)\bar{\sigma}$, and $d\bar{\epsilon}^p=G(2/c)(f)^{\frac{1}{2}}df$ is a quadratic form in $d\epsilon_{ij}{}^p$. In engineering notation Eqs. (32) when solved for the stress differences and shearing stresses become

$$\begin{aligned}\sigma_x-\sigma_y&=\frac{Bd\bar{\epsilon}_x{}^p-Cd\bar{\epsilon}_y{}^p}{Gdf(AB+BC+AC)};\ \text{etc.}\\\tau_{xy}&=\frac{1}{2}\frac{d\gamma_{xy}{}^p}{Gdf3D};\ \text{etc.}\end{aligned} \tag{34}$$

Substituting (34) in (31) gives f as a quadratic function of the strains divided by $(Gdf)^2$. The effective strain

[17] R. Hill, Proc. Roy. Soc. **A193**, 281–297 (1948).

increment is thus a constant times the square root of $f(Gdf)^2$. Noting that

$$(d\epsilon_x^p)^2+(d\epsilon_y^p)^2+(d\epsilon_z^p)^2=-2d\epsilon_x^p d\epsilon_y^p -2d\epsilon_y^p d\epsilon_z^p-2d\epsilon_z^p d\epsilon_x^p \quad (35)$$

gives

$$d\bar{\epsilon}^p=2G(f)^{\frac{1}{2}}df/c$$

$$=\frac{(2)^{\frac{1}{2}}}{c}\left\{\frac{B(d\epsilon_x^p)^2+C(d\epsilon_y^p)^2+A(d\epsilon_z^p)^2}{AB+BC+AC} +\frac{(d\gamma_{xy}^p)^2}{6D}+\frac{(d\gamma_{yz}^p)^2}{6E}+\frac{(d\gamma_{zx}^p)^2}{6F}\right\}^{\frac{1}{2}} \quad (36)$$

which reduces to the isotropic form (14) where all constants including c are unity. The constant $(2/c)$ was chosen so that

$$dW^p=\sigma_{ij}d\epsilon_{ij}^p=\sigma_{ij}G\frac{\partial f}{\partial\sigma_{ij}}df=G2fdf =\frac{2}{c}Gc(f)^{\frac{1}{2}}(f)^{\frac{1}{2}}df=\bar{\sigma}d\bar{\epsilon}^p.$$

Equation (36) can only be used if experiments verify the hypothesis that $G=G(f)$ or the plastic work done is path independent.

For completely general quadratic anisotropic loading functions the effective strain increment is found as in Eq. (36) to be the root of a quadratic in the plastic strain increments whose coefficients are simply related to the coefficients of the stresses in f. If f is written in terms of nine stress components

$$f=A_{ij}\sigma_i\sigma_j \quad (i,j=1,2,\cdots,9), \quad (37)$$

the stress-strain relations (8) become

$$d\epsilon_i^p=2GA_{ij}\sigma_j df \quad (38)$$

which may be solved for the stresses to give

$$\sigma_i=B_{ij}d\epsilon_j^p/2Gdf \quad (39)$$

where B_{ij} is the inverse of A_{ij}; i.e., $B_{ij}A_{jk}=\delta_{ik}$. As for Eqs. (34), (36), if the plastic volume change is zero, the inverse is not unique and a variety of equivalent forms is possible. Using the same definition of effective stress and strain as before, $\bar{\sigma}=c(f)^{\frac{1}{2}}$, $d\bar{\epsilon}=(2/c)G(f)^{\frac{1}{2}}df$, gives

$$d\bar{\epsilon}^p=(1/c)(A_{ij}B_{im}d\epsilon_m^pB_{jn}d\epsilon_n^p)^{\frac{1}{2}} =(1/c)(B_{mn}d\epsilon_m^pd\epsilon_n^p)^{\frac{1}{2}}. \quad (40)$$

If experiments show that the quadratic anisotropic f of Eq. (37) is not satisfactory for experimental correlation, loading functions of higher degree in the stresses may have to be used. As in the isotropic case, if the assumption that plastic work is path independent, $G=G(f)$, is dropped, the effective strain may well be defined by the second method in the same form (40) as for the quadratic case, simply take

$$d\epsilon_i^p\propto\frac{g(f)}{[B_{ij}(\partial f/\partial\sigma_i)(\partial f/\partial\sigma_j)]^{\frac{1}{2}}}\frac{\partial f}{\partial\sigma_i}df, \quad (41)$$

so that

$$d\bar{\epsilon}^p\propto(B_{ij}d\epsilon_i^pd\epsilon_j^p)^{\frac{1}{2}}=g(f)df. \quad (42)$$

Little more can be said about the proper forms of $\bar{\sigma}$ and $\bar{\epsilon}^p$ as experimental data is very meager. It is hoped that this type of approach will serve as a useful guide in future experimental research.

CONCLUSION

Definitions of effective stress $\bar{\sigma}$ and effective strain $\bar{\epsilon}^p$ which offer the possibility of better understanding of tests on isotropic and anisotropic work hardening materials are given on the basis of stress-strain relations from the recent "stress" theories of plasticity.

The effective stress $\bar{\sigma}$ is defined in terms of the loading function f associated with the given material. A comparison is made between two fundamentally different definitions of the effective strain $\bar{\epsilon}^p$. One, now in use, assumes that the work dissipated depends on the final state of stress alone and not on the path, and the other presented here analyzes the stress-strain relations. For $f=J_2$ the two resulting definitions of $\bar{\epsilon}^p$ coincide and, for certain other loading functions, no matter how complicated the path of loading, agree within practical limits. For many purposes, therefore, the definition of effective strain given by Dorn

$$\bar{\epsilon}^p=\int_{\epsilon^p}(d\epsilon_{ij}^pd\epsilon_{ij}^p)^{\frac{1}{2}}$$

is entirely satisfactory even though $f=f(J_2, J_3)$. Fundamental differences exist, however, and the basic work assumption must be further examined as pertinent experimental data is obtained.

Although the several $\bar{\epsilon}^p$ and $\bar{\sigma}$ as defined here will correlate a variety of tests in a useful manner, the concepts of effective strain and stress must break down for loading paths in which Bauschinger and similar effects are present.

Formation of Crystal Nuclei in Liquid Metals

D. TURNBULL
Research Laboratory, General Electric Company, Schenectady, New York
(Received April 13, 1950)

The known facts about nucleation phenomena in liquid metals are interpreted satisfactorily on the basis of the critical size and interfacial energy concepts. In large continuous masses nucleation is almost always catalyzed by extraneous interfaces. However, in very small droplets the probability that a catalytic inclusion is present is so much less that their minimum nucleation frequencies are reproducible and form a consistent set of values.

Interfacial energies, σ, between crystal nuclei and the corresponding liquids have been calculated from nucleation frequencies of small droplets on the basis of the theory of homogeneous nucleation. Energies of interfaces, σ_a, one atom thick and containing N atoms were calculated from the σ's. The ratio of σ_a to the gram atomic heat of fusion, ΔH_f, was approximately 0.45 for most metals but $\sim$0.32 for H_2O, Bi, Sb, and Ge.

The effect of relative complexity of crystal structure upon the supercooling behavior of pure metals apparently is a reflection of its effect upon ΔH_f.

THE solidification of pure metals may be thought of as a sequence of two steps: nucleation of crystals and their subsequent growth. It is known that the growth of large metal crystals into their supercooled melts is very rapid and usually limited only by the rate at which the heat of solidification can diffuse from the interface. On the other hand the time preceding the appearance of "macroscopic" crystals may be very long relative to the duration of rapid growth and often controls the solidification rate. This time delay is known as the "nucleation" period. The reciprocal of the average nucleation period is the nucleation frequency or rate and is proportional to the volume of liquid or to the area of interfaces in the system that catalyze nucleation.

Nucleation rates measured on large continuous liquid metal masses are not reproducible. However, recent experiments[1,2] on the supercooling of small metal droplets have shown that the temperatures at which the rate of crystal nucleation becomes appreciable in them are consistent and reproducible to within about ± 5 percent.

Generalizations about the solidification behavior of small droplets have been made in a preceding paper.[2] By combining these generalizations with the known facts of the supercooling behavior of large masses, a theory of nucleation can be constructed that accounts qualitatively for the nucleation behavior of liquid metal masses of all sizes and surface conditions.

CRYSTAL NUCLEATION AND ITS CATALYSIS

In nearly all theories of homogeneous (non-catalyzed) nucleation of crystals in liquids it is supposed that the nucleation period is the time necessary for a crystal of a certain critical size to form by thermal fluctuations. The radius of a critical size nucleus is directly proportional to the interfacial energy between the liquid and crystal and inversely proportional to the difference in free energy between liquid and crystal phases of infinite volume. The free energy of a crystal of critical size is decreased by fluctuations that either add or subtract atoms to it; consequently, any crystal that is smaller than the critical size will usually disappear and any that becomes larger will usually grow.

On the basis of these concepts and absolute rate theory, the following expression has been derived[3] for the homogeneous nucleation frequency of crystals in supercooled liquids:

$$I = A \exp[-K\sigma^3/(\Delta F_V)^2 kT], \qquad (1)$$

where $A = n(kT/h) \exp[-\Delta F_A/kT]$ and the other symbols are defined as follows: I = number of nuclei/sec. $\times$cm^3, n = number of atoms in the mass of liquid, K = a factor determined by the shape of the nucleus, σ = interfacial energy/cm^2 between liquid and crystal, ΔF_V = difference in free energy/cm^3 between crystal and liquid phases of infinite volume, and ΔF_A = free energy of activation for transporting an atom across the liquid-crystal interface.

In Eq. (1) three parameters σ, K, and ΔF_A are not known. However, experience indicates that ΔF_A is very small, at least for large crystals, and there is reason to believe that it is of the same magnitude as the activation energy for viscous flow, $\Delta F_A'$. If $\Delta F_A \approx \Delta F_A'$ then $\exp(-\Delta F_A/kT)$ is of the order of 10^{-2} at the solidification temperature for most metals. Therefore, it is expected that $A \approx 10^{33\pm1}$ sec.$^{-1}$ cm^{-3}. A can be evaluated from measurements of the homogeneous nucleation frequency as a function of temperature and its value compared with the predictions of theory. If the theory is thus approximately verified, Eq. (1) can be used to calculate σ (assuming some shape for the nucleus) from the nucleation frequency measured at a single temperature.

There is convincing evidence that nucleation of crystals in large continuous masses of liquid metals is almost always catalyzed by accidental inclusions and container walls.[4-6] The theory of nucleation catalysis is

[1] D. Turnbull, J. Metals **188**, 1144 (1950).
[2] D. Turnbull and R. E. Cech, J. App. Phys. **21**, 804 (1950).
[3] D. Turnbull and J. C. Fisher, J. Chem. Phys. **17**, 71 (1949).
[4] W. T. Richards, J. Am. Chem. Soc. **54**, 479 (1932).
[5] D. Turnbull, J. Chem. Phys. **18**, 198 (1950).
[6] J. H. Hollomon, A.S.M. Symposium on Thermodynamics, Cleveland (1949).

also based upon the critical size concept. When crystals form on the surfaces of catalytic interfaces the total increase in free energy due to forming the crystal surfaces is less than the increase that would accompany the formation of the same mass of crystal without the aid of extraneous interfaces. Catalysis requires that interfaces be present such that the equilibrium contact angle, θ, made by the crystal and the interface in the presence of liquid be less than 180°. An expression can be written for the frequency of heterogeneous (catalyzed) nucleation analogous to the expression for homogeneous nucleation as follows:[7]

$$I_s = A' \exp[-K\sigma^3\{f(\theta)\}/(\Delta F_V)^2 kT], \qquad (2)$$

where $A' = n_s(kT/h)\exp[-\Delta F_A/kT]$, $f(\theta) = (2+\cos\theta)\times(1-\cos\theta)^2/4$, I_s = nuclei/sec. cm^2, n_s = number of atoms in the interface/cm^2, and A' is estimated to be $\approx 10^{25\pm1}$ sec.$^{-1}$ cm^{-2}. Thus a clear distinction can be made between homogeneous and heterogeneous nucleation on the basis of the coefficient A evaluated from data on the nucleation frequency as a function of temperature. However, in large masses of liquid the nucleation frequency usually follows no regular relation such as (2) because accidental catalysts of varying degrees of effectiveness (i.e. different θ values) are liable to be present in different samples. Due to these difficulties adequate testing of the nucleation theory with data on large masses of liquid metals is almost impossible.

One approach to the problem of measuring the rate of homogeneous nucleation is to pass the liquid through fine filters prior to the rate studies in order to remove catalytic particles. It is known that the frequency of nucleation in filtered samples often is much smaller than in unfiltered ones.[8,9] However, the method is not promising for the investigation of nucleation in metal samples. Because of the relatively rapid growth rate of metal crystals, a nucleus originating at a single catalytic site can quickly transform a very large mass of liquid.

A method of eliminating the effects of nucleation catalysis that appears more promising for liquid metals is to break the sample into a number of non-communicating droplets that is large in comparison with the number of catalytic sites. This technique has recently been used to study the solidification behavior of a large number of pure metals.[1,2]

Data on the isothermal rates of solidification of small droplet aggregates of tin[10] and mercury[11] at several temperatures are now available. The temperature dependence of the nucleation frequency is so great that it was easily measurable only in narrow ranges of temperature corresponding to 59 to 63° supercooling for mercury and 100 to 115° for tin. Analysis of the data showed that $A = 10^{30.4}$ for tin and $10^{30.9}$ for mercury. Considering the theoretical and experimental uncertainties these numbers are in good agreement with the value for $A \approx 10^{33}$ calculated from the theory of homogeneous nucleation but considerably larger than the value 10^{25} to be expected if nucleation were catalyzed by a film coating the entire surface of all the droplets. Thus it appears that the conditions for homogeneous nucleation were closely approached in these experiments.

Because of the sharp temperature dependence of the nucleation frequency shown in these experiments (for mercury, I changed by a factor of 10 in 1½°C), it is possible to specify a narrow range, δ, of supercooling ΔT_- such that the nucleation frequency is practically zero when $\Delta T_- < \delta$ and immeasurably fast when $\Delta T_- > \delta$. On the basis of these results it is expected that crystals would be formed at a measurable rate in small droplets of other metals within a comparatively narrow temperature band characteristic of the metal. Microscopic observations of the solidification of small droplets (10 to 100 micron diameter) of many other metals confirmed this expectation.[2]

A maximum supercooling $(\Delta T_-)_{max}$ was found for each pure metal corresponding to a temperature at which the nucleation frequency becomes appreciable. Not all droplets supercool as much as $(\Delta T_-)_{max}$ but under suitable conditions a large fraction do. For most metals $(\Delta T_-)_{max}$ is of the order of 0.18 times the absolute melting temperature, T_0. Also, it is found that $(\Delta T_-)_{max}$ is reproducible (±5 percent) and not dependent upon the source of the metal in cases where this factor was varied.

SIZE EFFECT IN NUCLEATION

According to the critical size concept, the frequency of homogeneous nucleation of crystals, I_v, in droplets of volume v is

$$I_v = Iv, \qquad (3)$$

where I is given by Eq. (1). From the results on mercury and tin I_v may be calculated for any volume of liquid metal or alternatively the amount of supercooling (ΔT_-) at which the rate of nucleation becomes appreciable in droplets of a given size can be calculated as a function of the volume. For example, mercury droplets 4 microns in diameter were found to solidify[11] with a frequency of about 10^{-5} sec.$^{-1}$ at $\Delta T_- = 60$. From the rate equation it was calculated that nuclei should form with the same frequency in a mercury droplet 1 cm^3 in volume at a supercooling $\Delta T_- \sim 49°$. Since similar relations for the rate of nucleation are expected to be valid for other substances it is inferred that large continuous masses of liquid metals (of the order of 1 cm^3 in volume) *completely free of catalytic inclusions* and *not subjected to mechanical vibration* should supercool about

[7] D. Turnbull, A.S.M. Symposium on Thermodynamics, Cleveland (1949).
[8] J. Meyer and W. Pfaff, Zeits. f. Anorg. Chemie **217**, 257 (1934).
[9] V. I. Danilov and V. Neumark, J. Exp. Theor. Phys. (U.S.S.R.) **10**, 942 (1940).
[10] B. Vonnegut, J. Colloid Sci. **3**, 563 (1948).
[11] D. Turnbull, J. Chem. Phys. **18**, 768 (1950).

0.8 of the maximum supercooling observed in small droplets $(\Delta T_-)_{max}$, before solidification.

In agreement with this view, large continuous masses of liquid have been supercooled occasionally almost as much as small droplets. For example, comparatively large masses of gallium,[1] water,[12] and iron[13] have been supercooled more than 0.73 $(\Delta T_-)_{max}$.

The fact that continuous liquid masses 1 cm³ or more in volume rarely supercool more than ~ 0.05 $(\Delta T_-)_{max}$ is consistent with the viewpoint that the probability of finding accidental inclusions effective in promoting crystal nucleation is much greater in large than in small masses. This probability should increase either as the volume or surface area of the droplet so that, for example, the chances of finding one accidental inclusion in a 1 cc mass of mercury is about 10^{11} times greater than the chance of finding an inclusion in a 4 micron droplet separated from it. Thus, it appears that the known facts about the supercooling of small droplets and large masses of liquid metals are adequately accounted for on the basis of the critical size concept.

THERMAL HISTORY EFFECT IN SOLIDIFICATION

It has been fairly well established that the thermal history dependence of the nucleation frequency usually observed in the solidification of large continuous liquid masses is due to catalytic inclusions and container walls.[4,5] It follows that the nucleation frequency of small droplets should not be thermal history dependent provided that the catalysts responsible for the effect are in fact segregated in a small proportion of the droplets. This prediction has been verified by experiments on bismuth.[1] The nucleation frequency in large continuous masses of bismuth has been shown to be very dependent upon thermal history[14] but in small droplet aggregates it is independent of thermal history.[1]

From the facts that the nucleation frequency of lead crystals in sulfate-coated droplets, though increased by the catalytic effect of the film, is apparently not dependent on thermal history, it may be inferred that the presence of nucleation catalysts is a necessary but not a sufficient condition for thermal history dependence. On the basis of the micro-cavity theory of the effect,[5] the following conditions should be fulfilled in order for a thermal history effect to be observed:

1. Suitable microcavities be present in the catalytic surface. These microcavities retain small crystals above the melting temperature, but in order to be effective their diameter has to be very small (*ca.* 10^{-6} cm).
2. The substance must penetrate and fill the microcavities.
3. The contact angle θ made by the crystal with the catalytic surface when immersed in liquid should be less than 90°.

ALTERNATIVE INTERPRETATIONS OF THE SUPERCOOLING OF SMALL DROPLETS

In the foregoing it has been established that the known facts about nucleation of crystals in liquids can be explained satisfactorily on the basis of the "critical size theory." It remains to be considered whether any alternative hypotheses might explain the phenomena as well.

The basic fact to be accounted for is the long waiting period prior to the rapid growth of large crystals. Apart from the critical size hypothesis such a result might be explained if very small crystals, assumed to be stable, grow into supercooled melts at a rate many orders of magnitude less than do large crystals.

Two hypotheses that could lead to such an effect are thought worthy of serious consideration. First, it might be assumed that crystals need to have some type of imperfections in their surface in order to grow rapidly. Conceivably the probability that small crystals contain such imperfections might be much smaller than for

TABLE I. Summary of data on supercooling of small droplets.

Metal	T_0	Entropy of fusion $(\Delta H_f/T_0)$	$(\Delta T_-)_{max}$	Reference	$(\Delta T_-)_{max}/T_0$	$\frac{[(\Delta T_-)_{max}/T_0]^{\frac{1}{2}}}{\times[(T_0-[\Delta T_-]_{max}/\Delta H_f]^{\frac{1}{2}}}$
Mercury	234.3	2.38	58	11	0.247	0.268
Gallium	303	4.42	76	1	0.250	0.218
Tin	505.7	3.41	105	10	0.208	0.216
Bismuth	544	4.60	90	1, 2	0.166	0.154
Lead	600.7	2.04	80	1	0.133	0.196
Antimony	903	5.28	135	2	0.150	0.154
Aluminum	931.7	2.74	130	2	0.140	0.183
Germanium	1231.7	4.94	227	2	0.184	0.177
Silver	1233.7	2.19	227	2	0.184	0.232
Gold	1336	2.27	230	2	0.172	0.222
Copper	1356	2.29	236	2	0.174	0.222
Manganese	1493	2.31	308	2	0.206	0.243
Nickel	1725	2.43	319	2	0.185	0.224
Cobalt	1763	2.08	330	2	0.187	0.239
Iron	1803	1.97	295	2	0.164	0.224
Palladium	1828	2.25	332	2	0.182	0.223
Platinum	2043	2.30	370	15	0.181	0.226
Water	273.2	5.28	39	25	0.143	0.149

[12] R. Smith-Johannsen, Science **108**, 652 (1948).
[13] Bardenheur and Bleckman, Stahl u. Eisen **61**, 49 (1941).
[14] W. L. Webster, Proc. Roy. Soc. **140A**, 653 (1933).

TABLE II.* Interfacial energies between various crystal nuclei and the corresponding liquid calculated from frequency of nucleation in small droplets.

Metal	Crystal structure	Interfacial energy σ ergs/cm²	σ_g Cal./g atom	$\sigma_g/\Delta H_f$	σ_g/T_0
Mercury	Hexagonal	24.4	296	0.53	1.32
Water	Hexagonal	32.1	461	0.32	1.69
Gallium	Orthorhombic	55.9	581	0.436	1.91
Tin	Tetragonal	54.5	720	0.418	1.47
Bismuth	Rhombohedral	54.4	825	0.33	1.52
Lead	Face centered cubic	33.3	479	0.386	0.80
Antimony	Rhombohedral	101	1430	0.302	1.59
Aluminum	Face centered cubic	93	932	0.364	1.00
Germanium	Diamond	181	2120	0.348	1.71
Silver	Face centered cubic	126	1240	0.457	1.00
Gold	Face centered cubic	132	1320	0.436	0.99
Copper	Face centered cubic	177	1360	0.439	1.01
Manganese	Tetragonal	206	1660	0.480	1.11
Nickel	Face centered cubic	255	1860	0.444	1.08
Cobalt	Face centered cubic	234	1800	0.490	1.02
Iron	Body centered cubic	204	1580	0.445	0.88
Palladium	Face centered cubic	209	1850	0.450	1.01
Platinum	Face centered cubic	240	2140	0.455	1.05

* Heats of fusion and absolute melting points used in the computations summarized in Tables I and II are those recommended by K. K. Kelley ["Contributions to the Data on Theoretical Metallurgy. V. Heats of Fusion of Inorganic Substances," Bureau of Mines Bulletin No. 393 (1936)] with the exception of the heat of fusion of germanium (6100 cals/g atom) that was calculated by R. A. Oriani of this Laboratory from equilibrium data on the binary systems Ge-Au, Ge-Pb, Ge-Ag.

large crystals so that ΔF_A would be larger for the former. However, it is not easy to construct a model that would account for such a marked effect of imperfections upon growth rate unless the interfacial energy between liquid and crystal is quite large. Thus, dissociation of the imperfection and critical size hypotheses is difficult. A further difficulty with the imperfection hypothesis is that it does not account for the effect of specimen volume upon nucleation frequency.

The other hypothesis considered is that minute amounts of soluble impurities inhibit nucleation in some way. In the experiments on the solidification of small droplets[1,2] impurity concentrations were usually of the order of 0.001 to 0.1 atomic percent. It does not seem likely that such small concentrations of impurity could inhibit nucleation by causing the liquid-crystal interfacial energy to be increased. However, it is well known that minute quantities of impurity sometimes profoundly affect the macroscopic growth rate of crystals so it is conceivable that the growth of very small metal crystals in supercooled melts is practically stopped by such impurities.

Nevertheless, there are convincing arguments against this interpretation of nucleation during solidification. A mechanism whereby impurities effectively inhibit the growth of microcrystals but not macrocrystals is possible but complicated. In the small droplet experiments $(\Delta T_-)_{max}$ was found to be independent, within experimental error, of the source and purity (within limits) of particular metals. Also it is important that the $(\Delta T_-)_{max}$ values obtained by Turnbull and Cech[2] and Mendenhall and Ingersoll[15] on the metals common to both investigations are in close agreement. This consistency in the experimental data would not be expected if the rate

[15] C. E. Mendenhall and L. R. Ingersoll, Phil. Mag. **15**, 205 (1908).

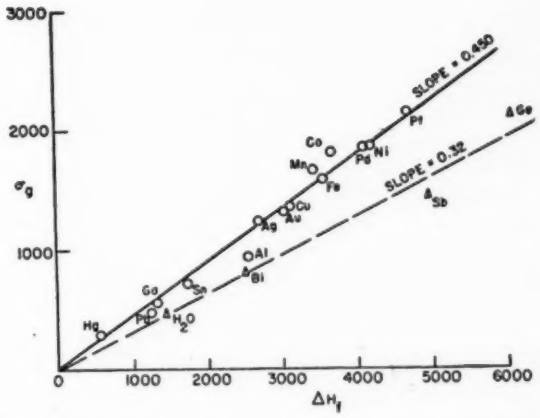


FIG. 1. Gram-atomic interfacial energy as a function of the gram-atomic heat of fusion for various substances.

controlling factor were inhibition by soluble impurities. Finally it appears that the wide difference in nucleation frequency for large and small masses of liquids cannot be explained on the basis of the inhibition hypothesis unless the critical size theory is also used. There is no reason why minor amounts of soluble impurities should be more effective in inhibiting crystal growth in small droplets than in large continuous liquid masses unless it is supposed that in the latter crystals are formed by a different mechanism such that their continued growth is not inhibited. It appears that the only other reasonable mechanism involves the catalytic action of insoluble impurities.

In view of these arguments it seems that the known facts about nucleation of crystals in liquids can be interpreted much more satisfactorily on the basis of the critical size theory than in terms of either of the slow-growth hypotheses examined. The excellent correlations of the small droplet supercooling data to be presented in the following sections constitute a further decisive argument in favor of the critical size interpretation.

CORRELATIONS OF DATA ON SUPERCOOLING OF SMALL DROPLETS

The ratio of the maximum supercooling observed in small droplets to the absolute melting temperature, $(\Delta T_-)_{max}/T_0$, is nearly a constant for many metals. These ratios for 50 micron diameter droplets of various metals are given in Table I. In calculating $(\Delta T_-)_{max}/T_0$ for iron and manganese there is the problem that the metals solidify in temperature ranges in which the most stable crystal phase is different from that in equilibrium with the liquid at the normal melting point. It is not known which phase nucleates most rapidly in the liquid at the solidification temperature but it was assumed arbitrarily that the solid phase stable at the normal melting point nucleates first. The assumption can cause little error in the iron calculation since the difference in free energy between the face centered and body centered cubic phases is very small.

For all but six of the substances $(\Delta T_-)_{max}/T_0$ is

≈ 0.185 with maximum deviations of ± 0.02. Considering all the data, the ratio ranges from 0.133 for lead to 0.250 for gallium.

When the implications of the approximate constancy of $(\Delta T_-)_{max}/T_0$ are examined in terms of the theory of homogeneous nucleation (see Eq. (1)) it is found that the ratio would be constant if σ were proportional to the gram atomic heat of fusion, ΔH_f, and if the entropy of fusion $(\Delta H_f/T_0)$ were constant. Entropies of fusion of metals are approximately 2.3 cal./deg.×g atom but there are some marked deviations from this value (e.g. 4.4 for Ga). These considerations suggested that a deeper insight into the phenomena might be gained by comparing σ directly with ΔH_f.

Values of σ were calculated from the nucleation frequency at one temperature and Eq. (1) with the aid of the following assumptions:

1. Crystal nuclei are spherical in shape so that $K = 16\pi/3$.
2. $\text{Exp}(-\Delta F_A/kT) = 10^{-2}$.
3. The entropy of fusion is independent of temperature.

The validity of these assumptions has been discussed elsewhere.[7] It follows from 3 that

$$\Delta F_V = \lambda \Delta T/T_0, \tag{4}$$

where λ = heat of fusion/cm^3 and $\Delta T = T - T_0$. It is believed that the available data are in general not sufficiently accurate to justify the use of a more precise free energy function. From the assumptions and Eq. (1)

$$\sigma = [\{3 \times 2.303 \lambda^2 (\Delta T)^2 kT/16\pi T_0^2\} \times \log(nkT \exp\{-\Delta F_A/kT\}/Ih)]^{\frac{1}{3}}. \tag{5}$$

I is estimated to be $10^{-(1\pm1)}$ sec.$^{-1}$ per 50-micron particle at the maximum supercooling. Although the uncertainty in I is very large, a factor of 10 error in its value introduces an error of only 1 percent in the calculated σ.

Since ΔH_f is a gram-atomic quantity it should be compared with a gram-atomic surface energy σ_g, which may be defined as the free energy of an interface containing Avogadro's number, N, atoms. If the area of such an interface is A,

$$\sigma_g = \sigma A. \tag{6}$$

Let it be assumed arbitrarily that the interface is one atom thick and let V be the gram-atomic volume; then,

$$A = N^{\frac{1}{3}} V^{\frac{1}{3}}, \tag{7}$$

where a structure factor of the order of unity is neglected. This treatment is analogous to that of Skapski[16] on liquid-vapor interfaces. Substitution of (7) into (6) gives:

$$\sigma_g = N^{\frac{1}{3}} V^{\frac{1}{3}} \sigma. \tag{8}$$

σ_g has been calculated for all of the substances whose solidification was studied by the small particle technique and the resulting values expressed in calories/gm atom are given in Table II.

Calculated ratios of σ_g to ΔH_f (see Table II) are fairly constant, ranging from about 0.31 for antimony to 0.53 for mercury. It is interesting that $\sigma_g \approx \frac{1}{2}\Delta H_f$ for one class of substances and $\approx \frac{1}{3}\Delta H_f$ for a smaller class. This difference is shown in Fig. 1 where σ_g is plotted against ΔH_f. The points for the more metallic substances (Class I) fall (within the experimental error) on a straight line of slope 0.45. Points corresponding to water and the semimetallic elements (antimony, bismuth, and germanium) (Class II) fall on a second line of slope 0.32.

Correct evaluation of σ_g depends upon the validity of the theory of homogeneous nucleation. There is an empirical correlation upon which the success of the correlation of σ_g with ΔH_f is based. This relation is:

$$[(\Delta T_-)_{max}/T_0]^{\frac{1}{3}}[\{T_0 - (\Delta T_-)_{max}\}/\Delta H_f]^{\frac{1}{3}} = M, \tag{9}$$

where M is constant for a class. M is equal to (see Table I) 0.222 ± 0.01 for Class I and 0.158 ± 0.01 for Class II.

The greater success of the $\sigma_g/\Delta H_f$ correlation relative to that of $(\Delta T_-)_{max}/T_0$ is in part illusory since the most important empirical factor (see Eq. (9)) leading to the success of the former is $[(\Delta T_-)_{max}/T_0]^{\frac{1}{3}}$. However, when this is corrected for, the $\sigma_g/\Delta H_f$ correlation is still the better of the two and in addition it appears to be more promising from a theoretical standpoint.

It is noted that the point for aluminum (see Fig. 1) on the σ_g *vs.* ΔH_f curve falls considerably below the Class I line on which it would seem to belong. Probably the most reasonable explanation of this disagreement is that surface films that may have catalyzed nucleation of aluminum crystals were not completely eliminated in the experiments.[2]

INTERPRETATION OF GRAM-ATOMIC INTERFACIAL ENERGIES

It is not surprising that mercury, tin, and the metals having cubical crystal structures fall into a distinct class with respect to solidification behavior. The four substances in Class II have the property of expanding upon solidification. In addition, their crystal structures are relatively more complex than the substances of I and excepting water over certain ranges of temperature they are all semiconductors. However, gallium exhibits all these properties yet it certainly falls in Class I. It is, of course, possible that the separation into two classes is more apparent than real. Perhaps the elements of Class II did not supercool the maximum amount for some unknown reason (in order to belong to Class I bismuth droplets should supercool 130° compared to the 90° observed) or approximations necessary in calculating σ_g may have caused misleading results. Nevertheless, the separation may be intrinsic for reasons not now apparent.

The parameter σ can be identified with the interfacial

[16] A. S. Skapski, J. Chem. Phys. **16**, 386 (1948).

energy between small crystal nuclei and liquid but it is important to know whether or not it corresponds to the interfacial energy between large crystals and their melts. Any such correspondence or lack of it can only be established by measurement of σ for large crystals. However, there is some indirect evidence that σ may not be markedly dependent on crystal size. The interfacial energies between liquid nuclei and vapor calculated from the nucleation rates[17, 18] measured by Volmer and Flood[19] are in good agreement with the values measured directly on extensive liquid surfaces. Apparently decreasing the size of liquid droplets to the dimensions of nuclei has little effect upon vapor-liquid interfacial energies.

Consider now the interfacial energy between adjacent crystals of widely different orientation σ_B (i.e., grain-boundary free energy). It follows that at equilibrium $\sigma_B \leqslant 2\sigma$ at the melting point of the substance. Fisher[20] has found σ_B between adjacent copper crystals of very different orientation to be about 535 ergs/cm^2 at 800°C. σ for copper solid-liquid is calculated from nucleation data to be 177 ergs/cm^2. Considering the experimental and theoretical uncertainties the numbers are not inconsistent with the condition that $\sigma_B \leqslant 2\sigma$.

It has been assumed in making the calculations that the nuclei are spherical i.e., that σ is independent of crystal orientation. There is some experimental justification of an indirect nature supporting this assumption. It has been found[21] that the microstructures of two phase lead-copper alloys equilibrated at relatively high temperatures can be interpreted satisfactorily by assuming that the liquid lead-solid copper interfacial energy is independent of the orientation of copper crystals. A number of other microstructures in two-phase alloys have been explained on the basis of analogous assumptions.

Recently it has been shown[22] that the interfacial energy between differently oriented adjacent crystals of silicon ferrite is practically independent of the orientation difference providing it is not very small or not of such magnitude that the orientations approach a twin relationship. When the orientations are very different it appears that the grain boundary should be highly disorganized so that the transition region between the grain boundary and adjacent crystal should be somewhat similar in nature to a liquid-crystal boundary. Thus, independence of grain boundary energy upon the orientation of the abutting crystals is a point in favor of the hypothesis that liquid-crystal interfacial energies for metals are isotropic.

[17] D. Turnbull and J. H. Hollomon, Sylvania Symposium "Physics of powder metallurgy," Bayside, Long Island (August, 1949).
[18] V. K. LaMer and G. M. Pound, J. Chem. Phys. **17**, 1337 (1949).
[19] M. Volmer and H. Flood, Zeits. f. physik. Chemie **170A**, 273 (1934).
[20] J. C. Fisher, private communication.
[21] C. S. Smith, Metals Tech. **15**, T.P. No. 2385 (1948).
[22] C. G. Dunn and F. Lionetti, J. Metals **1**, T.P. No. 2517 (1949).

Although the facts that have been presented do not prove that σ or σ_g values calculated from nucleation rates can be identified positively with the free energy of extended liquid-crystal interfaces it is believed that they afford substantial credibility for this identification.

EFFECT OF CRYSTAL STRUCTURE

It is well known that large masses of liquid metals that crystallize to close packed structures rarely can be supercooled more than 3 to 10 degrees while large masses that crystallize to more complex structures often can be supercooled much more. These facts have led to the concept[9] that the failure of large masses of the former class to supercool is an inherent property that is explained by the similarity in structure between close-packed crystals and their melts.

However, in view of the evidence that crystallization in large masses of all liquid metals is catalyzed by impurities, the facts are accounted for as well when it is assumed that effective catalysts for the nucleation of close-packed crystals occur much more frequently than those effective in the nucleation of crystals having more complicated structures.

Experiments on the supercooling of small droplets seem to have verified the latter theory. Small droplets of metals such as cobalt, silver, or nickel that crystallize to a cubical close-packed structure have been supercooled about as much relative to their melting points as have droplets of substances such as bismuth or tin that crystallize to more complex structures.

The arguments in favor of the concept that liquids crystallizing to close-packed structures should supercool much less than those crystallizing to more complicated structures hinge upon the idea that because of the structural similarity between the liquid and crystal[23] the interfacial energy between them should be small. This idea is quite reasonable, but by the same argument the heat of fusion, and therefore the free energy decrease in forming close-packed crystals should be and is small relative to that accompanying the formation of non-close-packed crystals.

Since it appears that σ_g is directly proportional to ΔH_f for most metals the effect of crystal structure upon supercooling behavior is but a reflection of its effect upon ΔH_f within one class of substances. However, additional insight into the effect of crystal structure upon σ_g can be gained by comparison of the σ_g/T_0 values listed in Table II. These ratios are roughly in the order of complexity of crystal structure and range from 1.91 for gallium to 0.80 for lead.

FURTHER IMPLICATIONS OF SMALL DROPLET RESULTS

The relation between σ_g and ΔH_f that has been found is analogous to similar relations between liquid-vapor

[23] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943), pp. 224–229.

interfacial energies and heats of vaporization.[16] It suggests that analogous relations exist for liquid-crystal interfacial energies of classes of substances other than those studied in these investigations and in other types of phase transformations. For example, the interfacial energy between two solid phases in solid-state transformations of a certain class might turn out to be proportional to the heat of transformation. Unfortunately, not much effort has been made to minimize possible nucleation catalysis in solid-solid reactions so it is not known to what extent the results now available have been affected by this factor. It may be possible in some instances to minimize this factor by applying the small particle technique to solid-state reactions.

Finally, it appears that one of the most promising fields of further research in phase transformations will be that of seeking what generalizations can be made about the chemical and structure relations between nucleation catalysts and the phase that is nucleated. Already considerable progress has been made in this direction by Vonnegut[24] and Schaefer[25] on the catalysis of ice nucleation.

ACKNOWLEDGMENT

The writer has benefited from conversations with Drs. J. C. Fisher and J. H. Hollomon on some of the subjects herein discussed. He has also been helped by the encouragement of Dr. J. H. Hollomon throughout the investigations on the supercooling of small droplets.

[24] B. Vonnegut, J. App. Phys. **18**, 593 (1947).
[25] V. J. Schaefer, Chem. Rev. **44**, 291 (1949).

Hot-Cathode Arcs in Cesium Vapor * †

RICHARD K. STEINBERG
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received April 21, 1950)

A hot-cathode arc in cesium vapor has been studied by the use of a plane Langmuir probe to investigate the potential distribution, electron temperature and electron density in the plasma. The range of arc currents and vapor pressures over which satisfactory probe measurements may be made are determined.

In some cases it was possible to maintain the discharge with a total arc drop of less than the ionization potential of cesium (3.87 volts). Potential "humps" along the arc are shown to be absent, and ionization by the high energy electrons of a Maxwellian electron velocity distribution is found unsatisfactory as a possible mechanism for positive ion production in the arc. The mechanism of successive electron collisions can account for the observed ionization on the assumption that the effective lifetime in the excited state is increased by the process of the imprisonment of resonance radiation.

I. INTRODUCTION

THE principal purpose of this work was to obtain information which would lead to a better understanding of the mechanism by which positive ions are produced in an electric arc in cesium vapor, particularly when the total voltage across the arc is less than the ionization potential of cesium. Although various explanations for low voltage arcs have been advanced, none of them seems to enjoy universal acceptance.

Of these explanations, at least four appear to be plausible. One is that the potential gradient is not uniform along the length of the arc, but instead, a potential "hump" may exist somewhere in the arc, and there may be two points or regions with a potential difference at least as great as the ionization potential. Another explanation involves the existence of oscillations in the arc. These oscillations may be of the relaxation type at frequencies of the order of kilocycles per second, or they may be, crudely speaking, bunches of electrons or positive ions vibrating in the electric fields produced by their own displacements from their equilibrium positions. The frequencies of the latter type of oscillations may be as high as thousands of megacycles per second. With oscillations in the arc, there is the possibility that potential differences greater than the ionization potential of the gas will be present even when the average total arc drop is much less. A third explanation is based on the fact that probe measurements on the plasma of an arc usually indicate that the velocity distribution of the electrons is reasonably close to Maxwellian, at least over the range of energies for which probe measurements can be made. Without attempting to give a detailed explanation of how the high electron energies are attained, this velocity distribution function is assumed to be obeyed for all energies. Then from the probability of ionization function of the gas under consideration, it is ascertained that the number of electrons with energies greater than the ionization potential is sufficient to create ions as fast as they diffuse from the plasma to the walls. The fourth explanation does not require that a single electron deliver the entire amount

* This work has been supported in part by the Signal Corps, the Air Materiel Command and the ONR.

† This paper is an abridged version of a thesis submitted in June, 1949 at the Massachusetts Institute of Technology, in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Physics. It has appeared as Technical Report No. 128, Research Laboratory of Electronics, Massachusetts Institute of Technology.

TABLE I. Cesium vapor pressure and atom density.

Temperature °C	Vapor pressure mm Hg	Atom density per cc
30	0.0000023	$(7.3)10^{10}$
100	0.00054	$(1.4)10^{13}$
150	0.0089	$(2.0)10^{14}$
200	0.079	$(1.6)10^{15}$
250	0.45	$(8.3)10^{15}$
300	1.9	$(3.2)10^{16}$

of energy necessary for ionization; rather, it suggests that it is possible for an electron to deliver only enough energy to excite an atom, with the remainder of the ionization energy coming from one or more successive collisions with other electrons. For reasons to be explained, this last mentioned mechanism is believed to be the one by which the cesium vapor arc is maintained.

Only a small amount of work on cesium vapor arcs has been reported. A series of papers by Mohler[1] is probably the most outstanding.

II. THE EXPERIMENTAL TUBES

Two tubes were constructed. The first, which will be referred to as tube No. 1, contained a plane nickel cathode and a plane tantalum anode, both 2.5 cm in diameter, mounted in a glass tube with an inside diameter of 5.3 cm. The cathode and anode were mounted on a glass and quartz assembly which could be moved along the axis of the tube by tilting the tube and shaking it slightly. The cathode-to-anode separation was 12.7 cm. Both the anode and cathode were constructed with built-in tungsten heaters capable of raising their temperatures to at least 1000°C. Low resistance, flexible, copper wire was used for making connections between the movable assembly and the lead-ins through the glass envelope. Actually, for most of the measurements, the cathode heater was turned off entirely because the arc itself caused sufficient heating to maintain the required thermionic emission.

A plane tantulum probe with a diameter of 0.25 cm was installed in tube No. 1 about 1 cm from the tube wall and with the plane of the probe parallel to the tube axis. The variation in position of the probe with respect to the arc was obtained by moving the cathode-anode assembly. A small glass tube was placed over the probe lead and the edge of the probe so that only the front surface of the probe "faced" the discharge. The same material was chosen for the anode and the probe to eliminate the uncertainty in the potential measurements which would otherwise arise from the contact difference of potential between different materials. In the case of cesium vapor, this difference would be particularly difficult to evaluate because cesium condenses on various surfaces to a thickness of the order of one atomic layer even when the temperature of the surface is hundreds of degrees centigrade higher than the temperature of the cesium supply. This layer of cesium lowers the work function of the surface by as much as 3 ev in some cases. The amount of lowering is strongly dependent on the fraction of a monatomic layer present and therefore dependent on the temperature.

Although annoying when making potential measurements, the cesium layer could be put to good use at the cathode. Thermionic currents of at least 2 amp. per cm² were attainable at cathode temperatures lower than 800°C because of the lowering of the work function of the cathode surface by the cesium. A text by Reimann[2] includes a more extensive discussion of the phenomenon together with references to original literature.

The cesium supply was contained in an appendix to the tube which consisted of a glass tube about 6 in. long. The cesium temperature, and therefore the vapor pressure, was controlled by the temperature of an oven which enclosed the appendix. To prevent bulk condensation of the cesium on the glass walls where it would cause low resistance leakage paths between the various

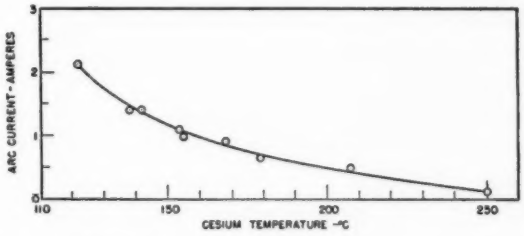


FIG. 1. Minimum arc current for non-oscillating condition as a function of cesium temperature (tube No. 1).

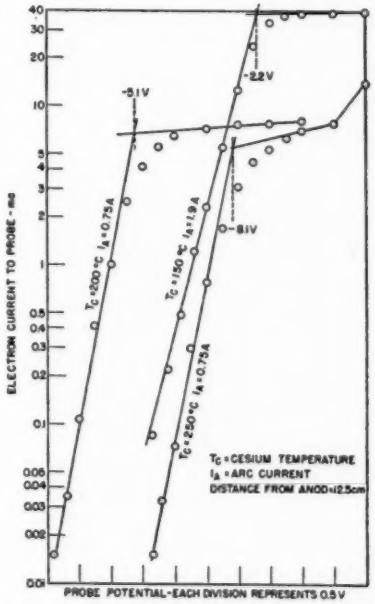


FIG. 2. Sample potential determinations (tube No. 2).

[1] F. L. Mohler, J. Research Nat. Bur. Stand. **9**, 25, 493 (1932); **10**, 771 (1933); **16**, 227 (1936); **17**, 849 (1936); **21**, 697 (1938). C. Boeckner and F. L. Mohler, J. Research Nat. Bur. Stand. **10**, 357 (1933).

[2] A. L. Reimann, *Thermionic Emission* (John Wiley and Sons, Inc., New York, 1934), p. 135 ff.

leads, a separate oven which could be maintained at a temperature higher than the temperature of the cesium was constructed for the main body of the tube. A glass window was installed in one side of the oven to provide a means for viewing the arc.

Tube No. 2 contained a series of probes in line as a means for measuring various plasma properties as functions of distance along the arc. In tube No. 2, the diameter of the front face of the cathode was 4.3 cm and the inside diameter of the glass tube was 4.6 cm. The anode was simply a disk 4.3 cm in diameter. The material for both the cathode and anode was tantalum. Four plane probes, each similar to the one used in tube No. 1, were mounted in a line along the length of the tube, and the distances from the center of the probes to the anode were 0.5, 2.5, 12.5, and 14.7 cm. The cathode-to-anode separation was 15.2 cm.

Both tubes were baked at 500°C for one hour, and the cathode and anode of each were heated to a bright red for several minutes either by means of their heaters or by radiofrequency induction heating. The pressure at the time the tubes were sealed off the vacuum system, as indicated by an ionization gauge, was about $(3)10^{-6}$ mm Hg. This reading corresponds roughly to the vapor pressure of cesium at room temperature and is the highest vacuum that could be expected.

III. CESIUM VAPOR PRESSURE AND ATOM DENSITY

The most recent and probably the most accurate vapor pressure data which were found were those of Taylor and Langmuir.[3] They measured the vapor pressure up to a temperature of 73°C by utilizing a method which makes use of the fact that practically every cesium atom which strikes a hot tungsten surface leaves as a positive ion. From their data, they derived the following formula for the vapor pressure of cesium in equilibrium with the liquid state:

$$\log_{10} p = 11.0531 - 1.35 \log_{10} T - (4041/T), \qquad (1)$$

where p is the pressure in millimeters of mercury, and T is the temperature of the cesium in °K. They claim an accuracy of 3 percent up to 327°C and 8 percent up to 727°C when extrapolating the curve beyond the range over which data were taken. Table I shows the cesium vapor pressure and atom density at various temperatures as determined from Eq. (1) and the equation of state.

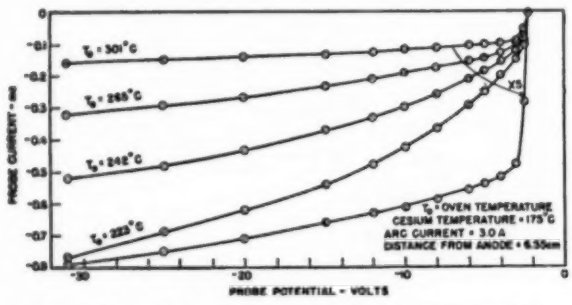


FIG. 3. Negative current portion of probe characteristic curves (tube No. 1).

[3] J. B. Taylor, I. Langmuir, Phys. Rev. **51**, 753 (1937).

IV. PLASMA POTENTIAL

The usual Langmuir technique was used for making measurements of the plasma potential. A text by Loeb[4] includes a discussion of the method and its limitations.

It was found that there was a rather limited range of arc currents over which reliable probe measurements could be made. Oscillations with frequencies in the range from 1000 to 10,000 c.p.s. would occur in the arc when the arc current was reduced below a certain critical minimum value. This minimum was a function of vapor pressure (cesium temperature) and was lower at the higher pressures, as shown in Fig. 1. The amplitude of these oscillations depended upon several parameters including vapor pressure, cathode temperature and arc current, but they usually had an amplitude of the order of 1 volt as indicated by an oscilloscope connected across the anode and cathode. The oscillations could not be eliminated by circuit modifications. Probe data were taken only at those values of arc current for which no oscillations existed.

The upper limit to the arc current at which reliable probe measurements could be made was set by the fact that at the higher arc currents the probe characteristic curve did not have a well-defined bend in it, or it had no bend at all, at plasma potential. Instead, the probe current would continue to rise as the probe voltage was increased until the probe was drawing the entire arc current. Also, the probe currents obtained at the higher arc currents were sufficient to cause probe heating, which, because of the cesium layer, changed the work function of the probe and introduced another source of error in the measurements. With tube No. 1, good probe characteristic curves were obtained with arc currents up to 3 amp. and up to 1.9 amp. with tube No. 2. The limitations just described were such that no reliable probe measurements at all could be made at cesium temperatures below about 150°C. Measurements were not made at cesium temperatures higher than 250°C because of possible damage to the tubes by chemical attack by cesium.

Sample logarithmic plots of probe characteristic curves are shown in Fig. 2.

Probe data taken at potentials negative with respect to floating potential were found to be of little value. Figure 3 shows the negative portion of the probe characteristic curves taken with the plane probe in tube No. 1 when the cesium temperature and arc current were held constant at 175°C and 3 amp. respectively, but when the main oven was at various temperatures from 223°C to 301°C. The lower oven temperatures appear to cause a much greater positive ion current to the probe. Similar sets of curves were obtained at cesium

[4] L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases*, (John Wiley and Sons, Inc., New York, 1939), pp. 232–257.

temperatures of 125°C and 250°C. It was discovered that even when the oven temperature was much higher than the cesium temperature, enough cesium would condense on the glass surfaces in the tube to cause leakage which, although small, was great enough to cause substantial errors in the measurement of positive ion currents. Thus, since the plane probe was doubtlessly in contact with its glass shield, the shield could have acted as a part of the probe, with the effect of increasing its area. The increase in leakage resistance with temperature would then account for the smaller probe currents found at the higher oven temperatures. The negative probe currents that were obtained at an oven temperature of 300°C were possibly a reasonably good measure of the positive ion current present, but, again because of the cesium layer on the surface, thermionic emission from the probe surface is of the same order of magnitude at this temperature.

From kinetic theory the relationship

$$I = n(kT/2\pi M)^{\frac{1}{2}} \qquad (2)$$

may be obtained, where I is the random current of gas particles crossing unit area in unit time, n is the number of particles per unit volume, k is Boltzmann's constant, T is the absolute temperature and M is the mass of the particle. This equation indicates that if the electron and positive ion densities and temperatures were equal, the ratio of random electron current to random positive ion current would be equal to the square root of the ratio of the ion to electron masses. The square root of this ratio is approximately 500 in the case of cesium. In order to determine plasma potential more accurately, the positive ion current was assumed to be 1/500 of the random electron current and this correction was applied to the measured values of probe current. The logarithmic plots so obtained (of which those in Fig. 2 are examples) were substantially linear in the region of rapidly varying electron current to the probe. The utilization of the data obtained at probe potentials more negative than floating potential was not attempted except to confirm, as near as could be ascertained, the fact that 500 was a reasonable value for the electron-ion current ratio.

Experimental work on mercury vapor by Robert Howe at the Massachusetts Institute of Technology, and by others, has indicated that the random electron-ion current ratio is actually less than the square root of the mass ratio in spite of the fact that the electron temperature is usually found to be quite high. In the plasma potential determination, a considerable variation in the ratio which is used causes only a small variation in the potential measurement.

TABLE II. Total arc drop (tube No. 1)

Cesium temperature °C	Arc current amp.	Arc drop volts
150	3.0	3.0– 9.0
200	3.0	6.6– 7.8
200	1.0	6.0–10.4
250	3.0	5.5–10.7
250	1.0	8.2–13.5
250	0.3	21.5

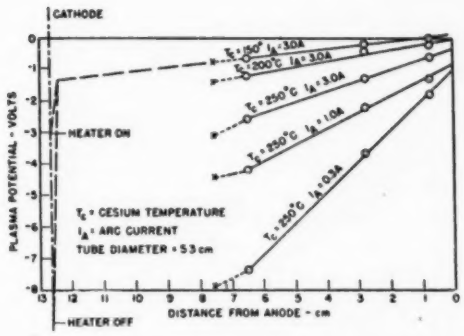


FIG. 4. Plasma potential (tube No. 1).

For tube No. 1 the plasma potential, referred to the anode as a function of distance from the anode, is shown plotted in Fig. 4 for various cesium temperatures and arc currents. It was intended to take data with the probe at positions close to the cathode but the movable assembly became damaged before this was accomplished. However it happened to be possible to take some data using the tantalum electrode as the cathode and the nickel electrode as the anode, with a cathode-to-anode separation of 8.5 cm and with the probe 0.9 cm from the cathode. The data so taken are represented by crosses and dotted lines in Fig. 4. The indication is that the plasma potential is at least a reasonably linear function of distance to within 0.9 cm from the cathode. Most of the data were taken with the cathode heater current turned off entirely and the arc itself was used as a source of cathode heating power. In order to obtain electrons for starting the arc it was necessary to use the cathode heater.

The total arc drop was a function of cathode temperature, and by using the heater, this total arc drop could be reduced to a value several volts below the value obtained without the heater. Excessive heating caused the cesium layer to evaporate faster than it was deposited, so that in spite of the increased temperature, the resulting increase in work function of the cathode surface caused a reduction in the available thermionic emission and an increase in arc drop. At a cesium temperature of 150°C and an arc current of 3 amp., for example, the total arc drop was 9 volts with the heater off. By using the heater, the drop could be reduced to a value as low as 3 volts. This example is shown graphically in Fig. 4.

The results, to be described later, obtained with tube No. 2 indicate that a linear extrapolation of the plasma potential to within 0.5 cm of the cathode is justified. Table II indicates the arc drops which were recorded at the times the probe data were taken.

With the tantalum cathode and nickel anode, the arc drop could be made as low as 2 volts. It was possible, though difficult, to reproduce these minimum values; the difficulty arose from the fact that it was necessary

that a large portion of the cathode be at the optimum temperature for thermionic emission and that the cathode heating by the arc tended to make the temperature non-uniform. The important point here, however, is that the potential of any point in the plasma when referred to the anode was found to be independent of cathode temperature and total arc drop. Several readings of potential were made under different conditions of arc drop and in each instance, even though the arc drop varied as much as 6 volts (the largest variation recorded), the potential of the plasma with respect to the anode stayed the same within experimental error, which was estimated to be about 0.1 volt. If the 150°C-3 amp. curve is considered again, it is seen that the negative anode fall is no greater than about 0.2 volt, and that with a total arc drop of only 3 volts there are no

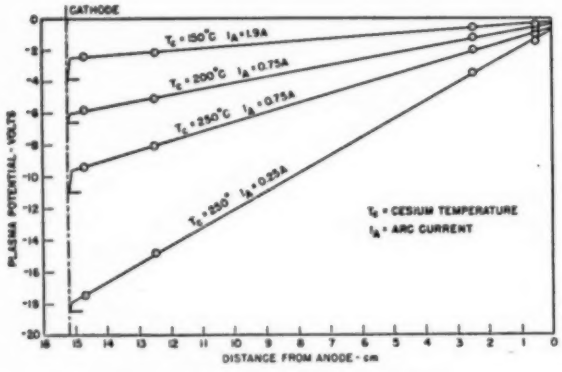

CATHODE
T_c = 150°C I_A = 1.9A
T_c = 200°C I_A = 0.75A
T_c = 250°C I_A = 0.75A
T_c = 250° I_A = 0.25A
T_c = CESIUM TEMPERATURE
I_A = ARC CURRENT
PLASMA POTENTIAL - VOLTS
DISTANCE FROM ANODE - cm

FIG. 5. Plasma potential (tube No. 2).

two points in the discharge with a potential difference as great as the ionization potential of cesium (3.87 volts).

The plasma potential measurements obtained from tube No. 2 are shown plotted in Fig. 5. It can be seen that they agree both qualitatively and quantitatively with those obtained from tube No. 1 if consideration is given to the fact that the cross-sectional area of tube No. 2 was only 0.75 times as great as that of No. 1.

Most of the data taken with tube No. 2 were taken with the cathode heater turned on and adjusted to yield minimum obtainable total arc drop. The total arc drop is indicated for each curve at the left edge of Fig. 5. The curve corresponding to a cesium temperature of 150°C and an arc current of 1.9 amp. shows no indication that there are two points in the tube with a potential difference greater than the ionization potential of cesium, a result similar to that obtained from tube No. 1. Cathode falls of from 0.8 to 1.5 volts were found.

The total arc drop as a function of arc current for tube No. 2 at the three vapor pressures studied is shown plotted in Fig. 6. For each curve the points represent the lowest drop which was attainable with the particular cathode which happened to be in use. For the higher arc currents, i.e., above about 3 amp., the heating of the cathode by the arc itself was sufficient to cause the cathode temperature to rise above optimum value for thermionic emission. Therefore it was necessary, in obtaining these data, to measure the arc drop immediately after starting the arc. Within a few seconds the arc drop would rise by as much as a volt or more above its initial value. If the cathode had been capable of emitting even greater thermionic currents the total arc drop might have been less because there was no apparent lower limit to the cathode fall other than zero.

The random electron current as a function of distance along the arc for two typical cases is shown in Fig. 7. The random current near the cathode was high, as shown, when the cathode heater was used to enable the cathode to emit a copious supply of thermionic electrons. When the heater was turned off, making it necessary for the arc itself to heat the cathode through the high cathode fall thus created, the random electron current near the cathode was as much as an order of magnitude less. This effect was in spite of the fact that no measurable difference in the space potential was noted.

An indication of the electron temperature may be obtained from the slope of the curve representing the logarithm of the electron current to the probe as a function of probe potential. Actually, the term "temperature" does not have a well-defined meaning when applied to the particles in the plasma of an arc; nevertheless, the temperature obtained in this manner is plotted as a function of distance along the arc in Fig. 8. Both of these sets of data were obtained with the four plane probes in tube No. 2. Similar results were obtained with tube No. 1.

V. OSCILLATIONS

Except for the low frequency oscillations, previously mentioned, which occurred when the arc current was reduced below a certain critical value, no oscillations were found. During all of the runs an oscilloscope which was responsive to frequencies up to 2 Mc was connected across the anode and cathode of the tube to check for the existence of oscillations, and data were taken only when no oscillations were indicated.

A search was made under various conditions of arc current and vapor pressure for oscillations at frequencies higher than might have been indicated on the oscilloscope. To carry out the search, a radio communications receiver, a radar search receiver and a crystal detector were employed in conjunction with various tuned circuits and Lecher wires connected in the arc circuit. A failure to find oscillations, no matter how thorough the search, does not establish the fact that no oscillations were present; however it is believed to have been reasonably well established that no oscillations were present at frequencies below an upper limit of about 10,000 Mc/sec. According to a derivation given by Tonks and Langmuir,[5] it is possible for electron oscillations in a plasma to occur at a frequency of $(8980)n^{\frac{1}{2}}$ where n is the number of electrons/cm³. The highest electron density found in the range of arc currents and

[5] L. Tonks, I. Langmuir, Phys. Rev. 33, 195, 990 (1929).

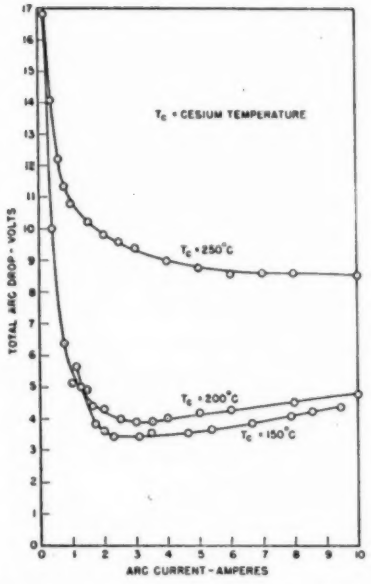


FIG. 6. Arc voltage *vs.* arc current (tube No. 2).

vapor pressures over which data were taken was $(1.6)10^{12}$ electrons/cm³. This density was obtained near the cathode of tube No. 2 when the cesium temperature was 150°C and the arc current was 1.9 amp., and it would imply a frequency of 12,000 Mc which might have been higher than the crystal was capable of detecting. However, 12,000 Mc is an upper limit to the possible frequency, and even under the same arc conditions the electron density near the anode was $(1.6)10^{11}$ electrons/cm³. This density would indicate a frequency of 3600 Mc, which was known to be within the range over which the detecting instruments would function.

High frequency oscillations have been reported occasionally in connection with tubes containing arcs, and certainly can be obtained with tubes of special design. The existence of the oscillations so found is usually not assured by the mere presence of the arc but is critically dependent on such factors as arc current and potentials of the various electrodes that may be present in the tube.

It is believed that there is little evidence to support a theory of ionization based on oscillations, particularly since, even if their existence were established with certainty, it would yet remain to be shown that they had sufficient amplitude to produce ionization. On the other hand, the present failure to detect oscillations is an indication that the ionization could not have been dependent upon them.

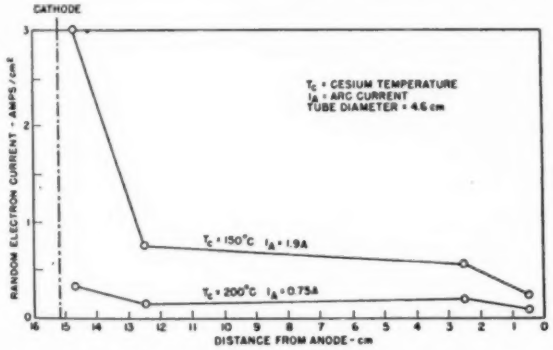

CATHODE
T_c = CESIUM TEMPERATURE
I_A = ARC CURRENT
TUBE DIAMETER = 4.6 cm
T_c = 150°C I_A = 1.9A
T_c = 200°C I_A = 0.75A
RANDOM ELECTRON CURRENT - AMPS/cm²
DISTANCE FROM ANODE - cm

FIG. 7. Random electron current density (tube No. 2).

VI. IONIZATION BY ELECTRONS WITH A MAXWELLIAN VELOCITY DISTRIBUTION

The number of ions lost to the walls per centimeter-length of tube may be obtained experimentally and is given by the expression $2\pi Ri_+/e$, where R is the radius of the tube in centimeters, i_+ is the positive ion current in amp./cm² as determined by probe measurement, and e is the electronic charge expressed in coulombs. For the range of atom density for which the discharge fills the whole tube, the number of ionizing collisions per electron per sec. is then

$$\alpha = 2\pi Ri_+/e\pi R^2 n = 2i_+/eRn, \qquad (3)$$

where n is the number of electrons per cm³, and, for the examples to be considered, it will be assumed that n is not strongly dependent on distance from the axis of the arc. If their velocity distribution is Maxwellian, the number of electrons per cm³ with velocities between v and $v+dv$ cm per sec is

$$nf(v)dv = 4\pi n\left(\frac{m}{2\pi kT_e}\right)^{\frac{3}{2}} v^2 \exp\left(-\frac{mv^2}{2kT_e}\right)dv, \qquad (4)$$

where m is the mass of the electron in grams, k is Boltzmann's constant expressed in ergs per degree, and T_e is the electron temperature in °K. The number of ionzing collisions per sec. that will be produced by an electron of velocity v is (PvN/N_0) where N is the number of atoms per cm³ present, N_0 is equal to $(3.5)10^{16}$ and is the number of atoms per cm³ when the temperature and pressure are 0°C and 1 mm Hg respectively, and P is the probability of ionization per centimeter-length of path when N_0 atoms per cm³ are present. Note that P is the product of the atomic cross section for ionization and N_0. The number of positive ions per cm³ produced per second by electrons with velocities between v and $v+dv$

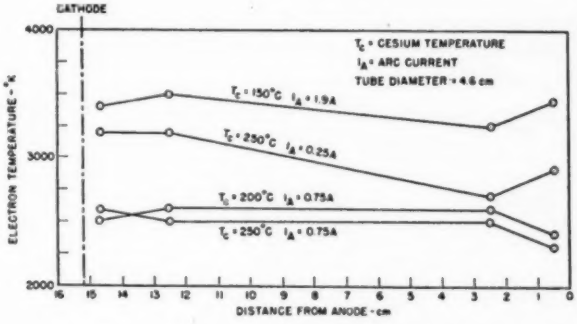

CATHODE
T_c = CESIUM TEMPERATURE
I_A = ARC CURRENT
TUBE DIAMETER = 4.6 cm
T_c = 150°C I_A = 1.9A
T_c = 250°C I_A = 0.25A
T_c = 200°C I_A = 0.75A
T_c = 250°C I_A = 0.75A
ELECTRON TEMPERATURE - °K
DISTANCE FROM ANODE - cm

FIG. 8. Electron temperature (tube No. 2).

is then

$$4\pi n\left(\frac{m}{2\pi kT_e}\right)^{\frac{1}{2}} P\frac{N}{N_0} v^3 \exp\left(-\frac{mv^2}{2kT_e}\right)dv. \quad (5)$$

If the substitution

$$v=(2Ve10^7/m)^{\frac{1}{2}} \quad (6)$$

where V is the electron energy expressed in electron volts, is made, expression (5) takes the form

$$(8)10^{14}\pi n\left(\frac{m}{2\pi kT_e}\right)^{\frac{1}{2}} \frac{e^2}{m^2} P\frac{N}{N_0} V \exp\left(-\frac{eV10^7}{kT_e}\right)dv. \quad (7)$$

The probability of ionization for cesium is not known, but it will be assumed for the purpose of this calculation that it is zero for electron energies less than the ionization potential V_i, and is constant with a value P_i, at all energies above V_i. After dividing by n, an integration of expression (7) from $V=V_i$ to $V=\infty$ yields the number of positive ions α' formed per second per electron. The result is

$$\alpha'=(2.0)10^{-7}\frac{P_iN}{T_e^{\frac{1}{2}}} \exp\left[-\frac{(1.16)10^4V_i}{T_e}\right] \times\left(V_i+\frac{T_e}{(1.16)10^4}\right), \quad (8)$$

where V_i is in volts.

By using the data obtained from tube No. 2 when the cesium temperature was 150°C and the arc current was 1.9 amp., for example, it was found that α, as computed from Eq. (3), was approximately 16,700 ionizing collisions per electron per sec. Although i_+ and n vary along the length of the plasma, they vary in approximately the same fashion so that α is reasonably constant. As described previously, i_+ was taken as 1/500 of the random electron current. Upon using 3.87 volts for V_i and 3400°K for T_e, it was found by equating α to α' that P_i must be at least 3000 ions per electron per cm of path. Data for the probability of collisions taken by Brode[6] indicate that it is 650 collisions per cm at V_i and drops off considerably at higher electron energies. Brode's probability of collision includes elastic and exciting collisions as well as ionizing collisions. From these results it is concluded that the number of electrons with sufficient energy for ionization which is made available from the Maxwellian distribution is not adequate to sustain the plasma. The results of additional examples are shown in Table III.

By working along similar lines Killian[7] concluded that, for mercury vapor, the ionization produced by this mechanism was sufficiently great in some cases to make up for the loss of ions caused by diffusion to the walls. However he obtained values for electron temperature up to 38,000°K, which is rather high. Because of the exponential term, α' is a rapidly varying function of temperature, and if the true temperature were even slightly less the mechanism would have been found insufficient in all cases.

[6] R. B. Brode, Phys. Rev. **34**, 673 (1929).
[7] T. J. Killian, Phys. Rev. **35**, 1238 (1930).

TABLE III. Data taken from tube No. 2.

Cesium temperature, °C	150	200	250	250
Atom density, per cc	$(2.0)10^{14}$	$(1.6)10^{15}$	$(8.3)10^{15}$	$(8.3)10^{14}$
Arc current, amp.	1.9	0.75	0.75	0.25
Electron temperature, °K	3400	2600	2500	3000
Electron density, per cc	$(4.8)10^{11}$	$(1.4)10^{11}$	$(9.5)10^{10}$	$(1.9)10^{10}$
Necessary P_i, per cm (see Section VI)	3000	18,000	66,000	380
Necessary P_s, per cm (see Section VII)	130	410	290	72

It should be noted that oscillations do not necessarily afford an explanation for the presence of high energy electrons or the existence of a high electron temperature, as is sometimes implied.[7,8] It is easy to see, though, why oscillations might cause probes to indicate, erroneously, high energy electrons or high electron temperatures. The explanation lies in the fact that, over a portion of the oscillation cycle, the probe at any potential setting is more positive with respect to the plasma than it would be in the absence of oscillations, and an increased electron current to the probe is produced. Because of the exponential current-probe potential relationship, the decrease in electron current that occurs over the opposite portion of the cycle does not cancel the effect. Thus at all probe potentials negative with respect to the average plasma potential, the electron current to the probe could be misleadingly large in the presence of oscillations.

An additional difficulty with the theory which assumes a Maxwellian velocity distribution of electrons lies in the inability to explain how the supply of high energy electrons is maintained in view of the fact that the number of high energy electrons would be expected to be diminished because of the inelastic (exciting and ionizing) collisions which occur. It is somewhat surprising that probe measurements indicate conformance to a Maxwellian distribution over as wide a range as they do, but this conformance is not a good reason to extrapolate the results beyond the range of measurements, which seldom extend to energies much higher than that corresponding to ionization potential. Further, measurements on the high energy electrons are subject to considerable error because of the uncertainties involved in extrapolating that portion of the probe characteristic curve which supposedly represents positive ion current.

VII. IONIZATION BY SUCCESSIVE COLLISIONS

The successive-collision hypothesis was suggested as far back as 1925,[9–11] but unless certain other phe-

[8] N. L. Oleson, C. G. Found, J. App. Phys. **20**, 416 (1949).
[9] K. T. Compton and C. Eckart, Phys. Rev. **25**, 139 (1925).
[10] W. Fabrikant, Techn. Phys. U.S.S.R. **5**, 864 (1938).
[11] B. Klarfeld, Techn. Phys. U.S.S.R. **5**, 913 (1938); Comptes Rendus de l'Acad. des Sciences U.S.S.R. **24**, 251 (1939).

nomena are taken into consideration, which was not always done, ionization by successive collision appears unlikely. The reason is that the lifetime of an excited state is so short compared with the average time between collisions that there is only a small probability that an excited atom will still be excited by the time a second electron strikes it. If the case of an arc current of 1.9 amp. and a cesium temperature of 150°C is taken as an example, and if Eq. (8), which assumes zero probability at energies below a certain critical value and a constant probability at higher energies, is employed, it is found that the number of excited atoms created per second per electron is $(8.1)10^3P_e$. The probability of excitation is P_e and the first critical potential of 1.48 volts was used. If $(4.8)10^{11}$ electrons per cm^3 is taken as the electron density and if the average lifetime of an excited state is 10^{-8} sec., it is found that the density of excited atoms is approximately $(3.9)10^7P_e$ per cm^3, which is about $(1.9)10^{-7}P_e$ times the density of unexcited atoms. Since an additional 2.39 ev are required to ionize an atom which has already been excited to the lowest excitation level, a second application of Eq. (8) indicates that the number of ions produced per second per electron is about $(1.1)10^{-4}P_eP_i'$, where P_i' is the probability of ionizing an excited atom if all the atoms were excited. In correspondence with the units used above, the atomic cross section for ionization of an excited atom is equal to P_i'/N_0. When the result just obtained is compared with α, the necessary number of ions per sec. per electron (16,700 for this example), it is seen that the product P_eP_i' must be at least $(1.6)10^8$ which is rather high. The assumption of a Maxwellian distribution of electron energies is possibly more justified in this case because the energies under consideration are not so great.

The phenomenon of imprisonment of resonance radiation is yet to be considered, and according to Holstein[12] the lifetime of an imprisoned light quanta in cesium vapor over a pressure range of 0.01–1 mm Hg is $(1.2)10^{-5}R^1$ sec., where R is the radius in centimeters of an infinitely long tube. Since it makes no difference which atom is excited, the passing of resonance radiation back and forth among the atoms has the effect of increasing the lifetime of an excited state by over 10^3 in the present case.

Also, according to Klarfeld,[11] the probability of ionization for excited atoms may be at least 10 times the probability for ionization from the ground state. The higher probability seems feasible when it is remembered that the wave function of an excited atom is much more spread-out than the wave function of an atom in the ground state. Thus when these two factors are considered it is found that P_e need not be greater than the square root of $(1.6)10^4$ or approximately 130 collisions per cm. This result causes ionization by successive collisions to appear more plausible. Additional examples are included in Table III.

VIII. CONCLUSION

Various properties of hot-cathode arcs in cesium vapor have been measured. In an effort to determine the mechanism by which the plasma of the arc is maintained it has been observed that no potential "humps" exist in the plasma, and, under certain conditions, no oscillations could be found. Sufficient ionization by the high energy electrons of a Maxwellian distribution appears unlikely. If the phenomenon of imprisonment of resonance radiation is taken into consideration, it has been shown that ionization by successive collisions is a plausible mechanism for plasma maintenance in spite of the short lifetimes of the excited states.

IX. ACKNOWLEDGMENT

I wish to thank Professor W. B. Nottingham, under whose direction the work was carried out, for the many helpful suggestions received. I also wish to thank Dr. A. W. Hull of the General Electric Company Research Laboratory, Schenectady, New York, for suggesting the problem and for assistance received in the preparatory phase of the work.

[12] Alpert, McCoubrey, and T. Holstein, Phys. Rev. **76**, 1257, 1259 (1949). (The information on cesium vapor was contained in a private communication from Holstein.)

Allotropy of Beryllium*

S. S. Sidhu and C. O. Henry
University of Pittsburgh, Pittsburgh, Pennsylvania
(Received April 24, 1950)

The allotropy of beryllium has been confirmed by employing different alloying properties of allotropic forms of metallic elements. Alloys of beryllium and gold, studied by x-ray diffraction methods, showed that the limit of solubility of gold in beryllium is between two and three atomic percent, and practically all of it is dissolved in the β-form. The lattice parameters of the two forms observed here are: $a_0=2.286$A, $C_0=3.588$A, and $C_0/a_0=1.570$ for the α-form, and $a_0=6.93$A, $C_0=11.35$A, and $C_0/a_0=1.638$ for the β-form. The volume of the unit cell of the β-form is about 29 times that of the α-form.

SOME investigators[1-3] have reported allotropy of beryllium while others[4] have failed to observe it. As a result a considerable uncertainty[5] exists regarding it. The object of this investigation was to obtain definite information in favor of or against such an existence.

EXPERIMENTAL PROCEDURE

The allotropic forms of metallic elements differ in their crystalline structures, and hence possess different alloying properties. These properties may be employed to prove or disapprove the allotropy of an element. If heavy atoms, such as of gold, can be dissolved, even to a limited solubility, in one of the allotropes of a light element as beryllium, noticeable changes may occur in the intensities of its diffraction lines due to a large difference in the atomic scattering factors of the two atoms. With this in view four alloys having compositions as given in Fig. 1 were prepared in an evacuated induction furnace with maximum pressure of the order of 10^{-4} mm Hg. Weighted mixtures of gold and beryllium metals were melted in BeO crucibles covered with BeO lids and enclosed in graphite heaters and ZrO_2 insulators. The maximum temperature varied between 1400° and 1480°C, and the solidification of the melts took place in the furnace. Powder x-ray diffraction patterns were made of stress relieved beryllium and representative samples of each alloy with a cylindrical camera of 114.6 mm diameter and copper $K\alpha$-radiation. The diffraction pattern of Be was repeated with a flat film and a specimen to film distance of 60 mm.

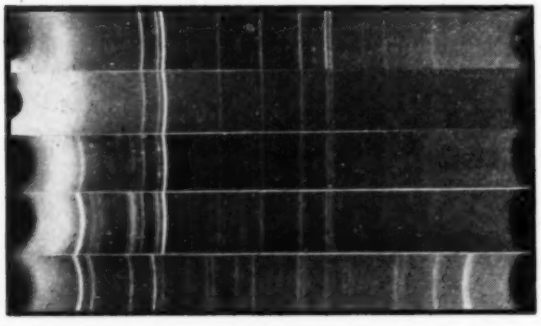


Fig. 1. X-ray diffraction photographs of: (A) α- and β-forms of Be; (B) Be with 0.02 atomic percent Au; (C) Be with 0.25 atomic percent Au; (D) Be with 2 atomic percent Au; (E) Be with 3 atomic percent Au. Camera diameter, 114.6 mm; Copper K_α-radiation.

RESULTS

As given in Table I the diffraction data obtained from the pattern of a sample of beryllium contained diffraction lines of the usual form of beryllium (α-form), and in addition several weak lines at relatively small Bragg angles. The interplanar spacings corresponding to the latter lines alone were not sufficient to determine a crystal structure but they could be fitted into an hexagonal structure with Miller indices as given under $\beta-$Be, and the lattice parameters as given below. These

Table I. Estimated relative intensities (I/I_0), interplanar spacings (d), and the Miller indices (hkl) for α- and β-forms of beryllium.

I/I_0*	$d_{(A)}$	$\frac{hkl}{(\alpha-\text{Be})}$	$\frac{hkl}{(\beta-\text{Be})}$
W	3.47	—	110
$V.W$	3.12	—	103
W	2.86	—	004
W	2.56	—	104
W	2.34	—	203
W	2.22	—	114
$V.W$	2.13	—	105
W	2.06	—	204
S	1.977	100	—
$V.W$	1.910	—	006
M	1.794	002	—
$V.S$	1.736	101	—
$V.W$	1.351	—	
M^+	1.328	102	—
$V.W$	1.272	—	
$V.W$	1.236	—	
S	1.143	110	—
S	1.023	103	—
W^+	0.991	200	—
S^-	0.965	112	—
M^+	0.955	201	—
$V.W$	0.911	—	
W	0.897	004	—
W	0.867	202	—
M	0.817	104	—

* V = Very; S = Strong; M = Medium; W = Weak.

* This work has been supported in part by ONR.
[1] F. M. Jaegar and J. E. Zanastra, Proc. Amst. Akad. Sci. **36**, 636 (1933).
[2] G. F. Kosolapov and A. K. Trapoznikov, J. Exper. Theor. Phys. U.S.S.R. **6**, 1136 (1936).
[3] G. P. Chatterjee and S. S. Sidhu, Phys. Rev. **76**, 175 (1949).
[4] P. Gordon, J. App. Phys. **20**, 908 (1949).
[5] L. Gold, AEC Declassified Report No. 2643 (July, 1949).

TABLE II. Estimated relative intensities and interplanar spacings for sample of Be with 2 atomic percent of gold.

I/I_0	d(A)
V. S	3.47
M⁻	3.03
M	2.81
W	2.45
V. W	2.33
S (double)	2.12
V. W	2.06
S	1.977*
V. W	1.910
S	1.794*
V. S	1.736*
V. W	1.639
W⁺	1.579
W	1.503
W⁺ (double)	1.376
M	1.328*
W	1.296
W	1.258
M (double)	1.222
V. W	1.173
S	1.144*
V. W	1.086
W	1.037
M	1.023*
V. W	0.990*
M	0.964*
M	0.954*
W	0.898*
W	0.865*
W	0.816*

* Lines of $\alpha-$Be.

TABLE III. Estimated relative intensities and interplanar spacings of sample of Be with 3 atomic percent gold.

I/I_0	$d_{(A)}$	$\frac{hkl}{AuBe_6}$	$\frac{hkl}{\alpha-Be}$
V. S	3.50	111	
S	3.03	200	
S	2.15	220	
W	1.977		100
S	1.831	311	
V. W	1.793		002
M	1.752	222	
M	1.733		101
W	1.520	400	
M	1.397	331	
M	1.362	420	
W	1.330		102
M	1.240	422	
M	1.173	333, (511)	
W	1.144		110
W	1.076	440	
M	1.030	531	
M	1.016	600	
V. W	0.989		200
V. W	0.965		112
M	0.958	620	
V. W	0.956		201
W	0.928	533	
W	0.919	622	
V. W	0.899		004
W	0.879	444	
V. W	0.869		202
M	0.853	551, (711)	
M	0.845	640	
M	0.815	642	104
M	0.794	533, (731)	

lines could not be attributed to impurities as determined by the spectrographic and x-ray diffraction analyses.

X-ray diffraction data obtained from patterns of alloys B, C, and D showed that there was no observable effect on the diffraction lines of the $\alpha-$Be with the increasing atomic percent of gold. The relative intensities and the interplanar spacings of these lines remained identically the same as those observed in the sample of beryllium. The diffraction lines of the $\beta-$Be on the other hand showed marked changes. The intensities of some of the original lines increased without any apparent change in their interplanar spacings, and additional lines appeared in increasing numbers as the atomic percent of gold increased. A few of the original lines appeared to be resolved into two lines. As given in Table II the diffraction data of sample D contained many more lines than those obtained from the sample of beryllium. These lines, however, were not due to gold or to the phase $AuBe_6$.

As shown in pattern *E*, Fig. 1, a definite transformation takes place at about 3 atomic percent of gold. The diffraction data, given in Table III, show conclusively that the alloy contained mostly $AuBe_6$ phase, which crystallizes in a face-centered cubic structure with $a_0=6.083$A, and a small percentage of $\alpha-$Be.

SUMMARY

The results obtained here showed that beryllium crystallizes in two hexagonal structures. The lattice parameters of the α-form are: $a_0=2.286$A, $C_0=3.588$A, and $C_0/a_0=1.570$, and of the β-form, $a_0=6.93$A, $C_0=11.35$A, and $C_0/a_0=1.638$. The volume of the unit cell of the β-form is about 29 times of the α-form. These parameters compare favorably with those reported by other investigators. The amount of crystalline material in the β-form appears to be very small and hence can be confused with impurities.

The limit of solubility of gold in beryllium is between 2 and 3 atomic percent, and practically all of it is dissolved in the β-form. Since this amount of gold in beryllium produces no observable effects in the α-form, but marked changes in the β-form, it would seem that the unit cell of one form is not related to that of the other and the two forms crystallize independent of each other.

Most of the prominent lines of the β-form, as seen in Table I, appear at relatively small Bragg angles; therefore, they are beyond the range of measurements made with the back-reflection method.[4]

The technique of employing different alloying properties of allotropic forms of metallic elements to prove their existence seems to work very well as demonstrated in this investigation.

The cooperation of Dr. Frank Foote and Mr. James Schumar of Argonne National Laboratory for getting alloys made for us is gratefully acknowledged.

Single Wave-Length X-Rays for Powder Diffraction

L. K. Frevel and P. P. North
The Dow Chemical Company, Midland, Michigan
(Received May 10, 1950)

WL_α radiation filtered through a Cu-Zn foil is essentially monochromatic and has the advantage over filtered CuK_α or MoK_α radiation in that the $WL_{\alpha 1}$ line is approximately ten times as intense as the $WL_{\alpha 2}$ line. Moreover the comparatively wide $\Delta\lambda$-spread ($WL_{\alpha 2}-WL_{\alpha 1}$) results in very sharp powder lines at those Bragg angles where the CuK_α doublet or MoK_α doublet is unresolved.

THE geometrical factors deciding the resolution of a Debye-Scherrer-Hull pattern are: R, radius of the camera; r, radius of the powder specimen; h, length of the specimen; δ, divergence of the impinging x-ray beam; and σ, difference in wave-length of the K_α doublet. For a fixed camera radius the optimum resolution corresponds to the limiting condition: $r, h, \delta, \sigma \rightarrow 0$. The geometrical resolution for a camera of radius R can be expressed approximately as the sum of δ, the divergence of the diffracted beam due to the divergence or convergence of the impinging primary beam, and $\omega(r, h)$, the angular spread of the diffracted beam due to the dimensions of the powder specimen [$\omega(r, 0) = 2r/R$ radian, for parallel x-rays].[1-3] Accordingly, two powder reflections d and $(d+\Delta d)$ will be resolved if

$$(\theta_{\alpha 2}-\theta_{\alpha 1})_d < (\theta_d-\theta_{d+\Delta d})_\alpha > (\delta+\omega)_\theta,$$

or

$$(\delta+\omega)_\theta < |(\theta_{d+\Delta d})_{\alpha 1}-(\theta_d)_{\alpha 2}|$$

where θ is the Bragg angle for the interplanar spacing d, and α denotes the wave-length under consideration. For example, with $MoK_{\alpha 1}$ radiation the two powder reflections corresponding to 2.260A and 2.274A will be separated by 3.4′, which interval is equal to the MoK_α doublet separation, $(\theta_{\alpha 2}-\theta_{\alpha 1})_{2.26}$. Evidently the use of very narrow slits does not aid greatly in the resolution of a non-cubic powder pattern because of the disturbing doublet separation. The utility and precision of powder diffraction work would be greatly enhanced if an intense source of strictly monochromatic $L_{\alpha 1}$ radiation could be made available (see Table I). Crystal-monochromatized $L_{\alpha 1}$ radiation should be quite excellent for single crystal diffraction techniques; however, for powder diffraction work such a source of monochromatic x-rays would be too weak for most analytical applications. To circumvent this limitation it was decided to use filtered L_α radiation. An examination of Table I reveals that tungsten offers the most convenient source of L_α radiation because of the general availability of x-ray tubes with tungsten targets and because of the absence of target contamination by sputtering or evaporation from

Table I. The L-series spectra and the K-absorption edges of filters for L_β and L_γ.

Z	$\alpha 1$	$\alpha 2$	$\beta 1$	$\beta 2$	$\gamma 1$	Z_f	K_∞
57 La	2.665A	2.674A	2.458A	2.302A	2.141A	22 Ti	2.496A
58 Ce	2.561	2.570	2.356	2.208	2.048	22 Ti	2.496
59 Pr	2.463	2.473	2.259	2.119	1.961	23 V	2.267
60 Nd	2.370	2.381	2.166	2.035	1.878	23 V	2.267
62 Sm	2.199	2.210	1.998	1.882	1.726	24 Cr	2.070
63 Eu	2.120	2.131	1.920	1.812	1.657	24 Cr	2.070
64 Gd	2.046	2.057	1.847	1.746	1.592	25 Mn	1.896
65 Tb	1.976	1.986	1.777	1.682	1.530	25 Mn	1.896
66 Dy	1.909	1.920	1.710	1.623	1.473	25 Mn 26 Fe	1.896 1.743
67 Ho	1.845	1.856	1.647	1.567	1.417	26 Fe	1.743
68 Er	1.785	1.796	1.587	1.514	1.364	26 Fe 27 Co	1.743 1.607
69 Tm	1.726	1.738	1.530	1.463	1.316	27 Co	1.607
70 Yb	1.671	1.682	1.476	1.416	1.268	27 Co 28 Ni	1.607 1.487
71 Lu	1.619	1.630	1.424	1.370	1.222	27 Co 28 Ni	1.607 1.487
72 Hf	1.569	1.580	1.374	1.327	1.179	28 Ni 29 Cu	1.487 1.380
73 Ta	1.522	1.533	1.327	1.285	1.138	28 Ni 29 Cu	1.487 1.380

Table I.—(*Cont.*)

Z	$\alpha 1$	$\alpha 2$	$\beta 1$	$\beta 2$	$\gamma 1$	Z_f	K_∞
74 W	1.476	1.487	1.282	1.245	1.098	29 Cu 30 Zn	1.380 1.284
75 Re	1.433	1.444	1.239	1.206	1.061	29 Cu 30 Zn	1.380 1.284
76 Os	1.392	1.402	1.197	1.171	1.025	29 Cu 30 Zn	1.380 1.284
77 Ir	1.351	1.363	1.153	1.135	0.991	30 Zn 31 Ga	1.284 1.192
78 Pt	1.313	1.325	1.120	1.102	0.958	30 Zn 31 Ga	1.284 1.192
79 Au	1.277	1.288	1.083	1.070	0.927	31 Ga 32 Ge	1.192 1.118
80 Hg	1.242	1.253	1.049	1.040	0.897	31 Ga 32 Ge	1.192 1.118
81 Tl	1.207	1.218	1.015	1.010	0.868	31 Ga 32 Ge 33 As	1.192 1.118 1.045
82 Pb	1.175	1.186	0.983	0.983	0.840	32 Ge 33 As	1.118 1.045
83 Bi	1.144	1.155	0.952	0.955	0.813	32 Ge 33 As 34 Se	1.118 1.045 0.980
90 Th	0.956	0.968	0.765	0.794	0.653	35 Br 36 Kr 37 Rb	0.920 0.866 0.816
92 U	0.911	0.922	0.720	0.755	0.615	36 Kr 37 Rb 38 Sr	0.866 0.816 0.770

[1] M. J. Buerger, *X-Ray Crystallography* (John Wiley and Sons, Inc., New York, 1942), pp. 397–434.
[2] *International Tables for Determination of Crystal Structures* (Gebrüder Bornträger, Berlin, 1935), Vol. II, pp. 581–4.
[3] H. Lipson and A. J. C. Wilson, J. Sci. Inst. 18, 144 (1941).

TABLE II. Comparison of powder diffraction patterns of $BaSO_4$.

MoKα	CuKα	d $WL_{\alpha 1}$	$WL_{\alpha 2}$	{hkl}
4.444A	4.418A	— A	A	200
4.332	4.324	—		011
3.893	3.905	3.917		111
3.787	3.775			201
3.580	3.572	3.586		002
3.443	3.449	3.452		210
3.316	3.320	3.322	3.322	102
3.104	3.104	3.106		211
2.836	2.835	2.836		112
2.725	2.729	2.730		020
2.485	2.481	2.482		212
2.322	2.326	2.321		220
		2.302		103
	2.280	2.280		302
2.209	2.214	2.207		221
	2.123	2.120		113
	2.110	2.104	2.102	312, 122
	2.058	2.054		410
	1.926	1.927		321
	1.860	1.859		303
	1.791	1.789		004
	1.760	1.758		031, 313, 123
	1.728	1.727		501
		1.681*		230
	1.678	1.679		421
	1.640	1.639		231
	1.594	1.596		502
	1.534	1.536		323
		1.528*		304, 512
	1.494	1.4954		024
	1.477	1.4751		314, 124
	1.459	1.4567		521
		1.4275*		503
	1.425	1.4234		332
	1.405	1.4065		430
		1.4020*		611
	1.383	1.3839		015
		1.3798*		431, 513
	1.365	1.3624		040, 205
	1.352	1.3492		414
		1.3260*		141
	1.323	1.3206		215
	1.301	1.2996		240, 333
	1.263	1.2619		523, 504
		1.2527		315, 125
		1.2267		514
		1.2186		711, 341, 225
		1.2027		405
		1.1953		532
		1.1472		630, 441

Radiation:	MoKα =0.7107A	CuKα =1.5418A	$WL_{\alpha 1}$ =1.4763A
Camera radius:	207.3 mm	71.7 mm	88.8 mm
Slit dimensions:	0.15×5×50 mm	0.4 mm rad.×70 mm	0.1×3×50 mm
Specimen radius:	0.2 mm	Wedge method	0.4 mm

For very absorbing materials such as $BaSO_4$, the wedge technique with CuKα gives very sharp lines because only the topmost part of the wedge is effective for diffraction. Pyrex capillary tubing was used to hold the powder sample for the MoKα pattern, and polystyrene capillary for $WL_{\alpha 1}$. Since the sample radius was excessively large in the latter case, only the top-half of the sample was bathed in the x-ray beam, this arrangement being comparable to the wedge technique.

* Resolved reflections.

the tungsten filament. A brass filter, 0.025 mm thick, was used to absorb the WL_β and WL_γ lines. The WL_α doublet thus transmitted has a decided advantage over filtered K_α doublets because the $WL_{\alpha 2}$ line is only one-tenth as intense as the $WL_{\alpha 1}$ line; so that for all practical purposes powder patterns taken with this filtered radiation will correspond to patterns taken with $WL_{\alpha 1}$. The comparatively wide $\Delta\lambda$-spread ($WL_{\alpha 2}-WL_{\alpha 1}=0.0111$A) has the further advantage of registering very sharp powder lines at those Bragg angles where the CuKα doublet or MoKα doublet is unresolved.

TABLE III. Powder diffraction data of Ag_2HgI_4.

CuKα d	I/I_1	$WL_{\alpha 1}$ d	I/I_1	$WL_{\alpha 2}$ d	I/I_1	{hkl}
6.273A	0.02	— A	—	A		200
5.635	0.03	—	—			210*
4.474	0.03	—	—			220
		4.198	0.03			300, 221*
		3.871	0.02			Phase II
3.646	1.00	3.647	1.00			222
3.502	0.02	3.497	0.10			320*
2.823	0.03	2.825	0.02			420
2.756	0.04	2.763	0.01			421*
		2.645	0.01			Phase II
2.570	0.02	2.576	0.10			422
2.475	0.01					510, 431*
		2.366	0.03			520, 432*
2.343	0.02	2.348	0.03			Phase II
2.231	0.50	2.233	0.70	2.234	0.01	440
		2.191	0.01			522, 441*
		2.137	0.30			531*
2.097	0.02	2.102	0.01			600, 442
2.076	0.02	2.079	0.01			610*
		2.015	0.01			Phase II
1.999	0.01	1.999	0.01			620
1.904	0.30	1.904	0.70			622
1.883	0.01					630, 543*
		1.823	0.10			444
1.732	0.01					720, 641*
1.686	0.02	1.686	0.01			642
1.616	0.02	1.617	0.02			650, 643*
1.576	0.04	1.578	0.06			800
1.530	0.01	1.532	0.01			820, 644
1.521	0.01	1.521	0.01			821, 742*
		1.509	0.01			653*
1.485	0.01	1.486	0.01			822, 660
1.449	0.10	1.449	0.30			662
1.440	0.01	1.440	0.02			832, 654*
		1.386	0.02			911, 753*
1.378	0.01	1.380	0.01			842
1.347	0.01					664
		1.367	0.01			920, 760*
1.310	0.01					852*
1.288	0.10	1.289	0.30			844
1.257	0.01					10·1·0, 942, 861, 742*
1.236	0.01	1.234	0.02			10·2·0, 862
1.214	0.04	1.216	0.03			10·2·2, 666

Legend: Filtered CuKα and WL_α x-rays were used to obtain the powder diffraction patterns of Ag_2HgI_4. {hkl} are the indices of the various reflections based on the unit cell $a=c=12.62\pm0.02$A. The complex iodide Ag_2HgI_4 was prepared by the dropwise addition of $AgNO_3$ solution to an agitated solution of K_2HgI_4. Chemical analysis of the washed and vacuum-dried precipitate showed 23.51 percent Ag (23.35 percent theor.), 53.25 percent I (54.94 percent theor.), and 20.52 percent Hg (21.71 percent theor.). The sample for the WL_α exposure had been aged for five years in a closed vial.

* Reflections requiring larger unit cell than $a=c=6.35$A.

Table II compares the powder patterns of $BaSO_4$ obtained with filtered CuKα, MoKα, and WL_α. It will be noted that six reflections were resolved with $WL_{\alpha 1}$ that appeared as unresolved lines with CuKα. Only two reflections out of fifty were attributable to $WL_{\alpha 2}$. For indexing non-cubic substances or reducing the superposition of powder lines in complex mixtures the use of $WL_{\alpha 1}$ radiation is to be recommended. To minimize continuous background radiation the tungsten x-ray tube with line focus should be operated at ~20 kv and ~40 ma.

The complex iodide, Ag_2HgI_4, is reported by Ketelaar[4] to be tetragonal with $a=c=6.34kX$. Patterns were taken of this substance with CuKα and $WL_{\alpha 1}$ radiation to ascertain if any splitting could be detected for pyramid reflections. Whereas no splitting was revealed (see Table III), a sufficiently large number of weak reflections were observed to require a larger unit cell. Extensive grinding of the complex iodide diffuses these additional sharp reflections into faint broad bands. It is quite likely that one is dealing with an order-disorder system in Ag_2HgI_4 and that the $E3_1$ structure will have to be reexamined in the light of the present findings.

[4] J. A. A. Ketelaar, Zeits. f. Krist. **80**, 190–203 (1931).

Shielding Properties of the Concrete Wall of the M.I.T. Cyclotron

VICTOR DELANO* AND CLARK GOODMAN
Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received May 15, 1950)

The shielding properties of the four-foot concrete wall of the M.I.T. cyclotron have been studied, using foils for neutrons and x-ray films for gammas. Fast neutrons were measured by counting the beta-active products of $Al^{27}(n, p)Mg^{27}$, $Fe^{56}(n, p)Mn^{56}$, and $Al^{27}(n, \alpha)Na^{24}$ reactions, thermal activation being eliminated by Cd holders. Resonance (1.44 ev) and thermal neutron distributions, measured by Cd-In-Cd and In, with suitable corrections, were reduced to absolute intensities by comparison with a known flux. Measurements were made with cylindrical concrete blocks fitted into a $4\frac{1}{2}''$ diameter hole which extended through the concrete wall. The results indicate an attenuation factor of 8.3 per foot for the fast neutron component and 10 per foot for thermal neutrons. An estimate of the fast neutron flux above 4.5 Mev showed a transmitted intensity slightly higher than the presently accepted tolerance value when bombarding Be. The absolute thermal neutron flux was below the tolerance level. The r.m.s. distance and the diffusion length for thermal and resonance neutrons in concrete were estimated by considering the distributed source to be equivalent to a point source at the inner face of the wall. The gamma-ray measurements were not convertible to roentgens. A mean energy of 2.7 Mev was obtained for the gamma-radiation within the wall. The attenuation factor for total gamma-radiation in the wall at large distances from the inner face was 5.9 per foot.

INTRODUCTION

THE occurrence of radiation cataract among the Japanese survivors at Hiroshima and Nagasaki and among cyclotron workers in this country has emphasized the importance of adequate neutron shielding. The problem is complicated by the invariable presence of gamma-rays which arise both from the reactions producing the neutrons and from the absorption and scattering of neutrons within the shield. The latter effect imposes important limitations on the choice of shielding material. What is needed is a combination of heavy and light atoms which are effective for attenuating gamma-rays and neutrons respectively. Of the many materials and combinations of materials investigated in the search for an ideal shield, concrete is one of the most satisfactory especially when cost is a major consideration and where size and bulk are not of too great concern.

While concrete has found wide use for the shielding of particle accelerators, there is only a limited amount of published information regarding its effectiveness.[1] Since the radiations from the M.I.T. cyclotron are representative of a number of existing sources and of others being built, the results of the tests made on this four-foot concrete shield may be of general interest.

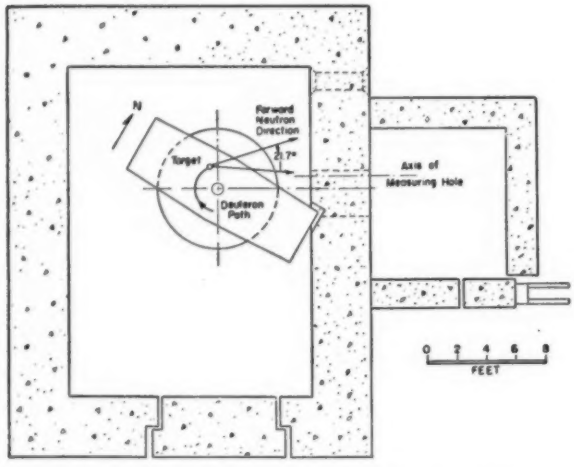


FIG. 1. Plan view of cyclotron shield, showing location of measuring hole.

EXPERIMENTAL ARRANGEMENTS AND PROCEDURES

Description of Cyclotron Shield and Measuring Hole

The M.I.T. cyclotron is shielded by a four-foot concrete wall topped with a three-foot concrete roof. Through the northeast wall of the shield, of which a plan view is shown in Fig. 1, a rectangular opening three feet wide and one foot high is provided for emergent beam experiments. When this beam is not in use, the opening can be closed with an array of concrete blocks of three sizes: $6''\times6''\times11\frac{1}{2}''$, $3''\times6''\times11\frac{1}{2}''$ and $6''\times6''\times6''$. These sizes are approximate because of slight surface irregularities.

Because of these irregularities and the relatively large dimensions, another set of blocks was constructed which formed a smooth, horizontal hole through the wall into which one could slide close-fitting cylindrical blocks for intensity measurements with foils and films. Since the center of the rectangular opening is below the central plane of the cyclotron dees, these blocks were designed to fit in the upper left-hand corner of the rectangular opening (see Fig. 2). By this arrangement the measuring foils could be exposed to the most energetic portion of the neutron beam.

As seen in Fig. 1, the axis of the measuring hole was displaced from the direction of the deuteron beam at the cyclotron target. For 15-Mev deuterons on Be, the maximum energy of neutrons to which the measuring hole was exposed was calculated as 18.8 Mev.

* Lieutenant Commander, U. S. Navy, now at Los Alamos Scientific Laboratory.

[1] M. S. Livingston, J. App. Phys. **15**, 2–19, 128–147 (1944); E. Creutz and K. Downes, J. App. Phys. **20**, 1236–1240 (1949); P. C. Gugelot and M. G. White, Phys. Rev. **74**, 1215 (1948); P. Jensen and O. Ritter, Zeits. f. Naturforschung **2a**, 376–384 (1947).

As shown in Fig. 3, three C blocks formed a base along the bottom of the rectangular opening for the A and B blocks. The measuring hole was formed using two pairs of A blocks, placed end to end. Two B blocks were set alongside these (see Fig. 2). All blocks were arranged to eliminate, as completely as possible, any continuous cracks or slits through the opening. Because of the small dimensional discrepancies in the latter, it was necessary to use wooden wedges between the blocks and the left-hand edge of the rectangular opening in order to force the hole-forming blocks tightly against the cyclotron wall. The original rectangular blocks were installed with their surfaces flush with the cyclotron wall in the cyclotron vault but they failed to complete the rectangular opening by about three-quarters of an inch along the outer side of the wall.

Each cylindrical D block has a $\frac{3}{16}''$ recess in one end in which were placed foils for neutron measurements or films for gamma-radiation measurements. By using two different lengths (2″ or 4″), measurements could be taken at two-inch intervals through the wall. "A" runs, as they are henceforth called, used twelve four-inch blocks. "B" runs used one two-inch block followed by eleven four-inch blocks and, at the end of the hole, another two-inch block. All the four-inch blocks were numbered and the two-inch blocks lettered to insure that successive "A" and "B" runs would have the same sequence of blocks.

The cylindrical blocks were always inserted with their recessed ends toward the cyclotron. The blocks were inserted by hand and removed with piano wire attached in a manner allowing half the blocks to be removed from the front of the hole and half from the back. Clearance between the top of the cylindrical blocks and the roof of the hole was $\frac{1}{16}''$ or less. Flake graphite was used as a lubricant.

Composition of Concrete

All the blocks used in the rectangular opening of the cyclotron wall were made from a concrete mix consisting of 1 part cement, $1\frac{1}{2}$ parts sand and 2 parts gravel (pea stone, $\frac{1}{2}''$ gravel), with about 5–$5\frac{1}{2}$ gallons of water per 100 lb. bag of cement. Composition of the cyclotron wall was slightly different, being made from a mix of 1 part cement, 2 parts sand and 4 parts gravel (1″ grated to $\frac{1}{4}''$), by weight. The difference in the mixes was necessitated by the relatively small size of the blocks which required more cement to insure adequate strength and proper shape. Table I gives chemical analyses of the two cements. Table I also includes the assumed composition of the cement used in the work of Jensen and Ritter.[2] The composition of the sand and gravel given in Table I is based on typical analyses of the sand and gravel from glacially deposited pits near Scituate, Massachusetts. Jensen and Ritter considered their sand and gravel combination to be pure SiO_2.

[2] P. Jensen and O. Ritter, see reference 1.

FIG. 2. Photograph of measuring hole showing insertion of foil and film-holding blocks.

An analysis of a sample block very kindly made by the Oak Ridge National Laboratory gave an average water content of 7.6 percent by weight. Using this value we have calculated the weight percentages of elements present as given in Table II.

The average density of the concrete in the cyclotron wall was measured as 2.20 g/cc, slightly less than Jensen and Ritter's value of 2.26 g/cc.

Counters and Counting Circuits

The beta-activities of the foils were measured with a Victoreen, mica-window, counter (Model VG-Special) mounted in a $1\frac{1}{2}''$ lead shield. The end-window of the counter was 2.0 mg/cm²; the foils were inserted on brass trays with a $\frac{1}{8}''$ deep recess, 3.18 cm in diameter centered directly below the counter window. The distance from the bottom of the tray to the center of the window was 1.0 cm. Circuitry consisted of a one-stage amplifier, a discriminator stage with adjustable bias and a scale-of-64, followed by a mechanical register.

For standardizing the counting rates, a brass slide was provided containing a small amount of RaD+E covered with 6 mils of aluminum which allowed only the 1.17-Mev beta-rays of RaE to emerge. The counter was shown to be linear up to at least 20,000 counts per

TABLE I. Composition of concrete mix used in cyclotron wall.

	Cement			Sand and gravel	
	Cambridge Cement Stone Company (movable blocks)	Boston Sand and Gravel Company	Cement used by Jensen and Ritter	Sand	Gravel
CaO	63.41%	63.3% (est.)	65%	0.25%	1.08%
SiO_2	20.03	22.2	25	95.87	73.60
Al_2O_3	7.10	5.1	7	1.83	14.44
Fe_2O_3	3.14	4.9	3	0.04	0.43
MgO	2.84	1.4			
SO_3	1.88	1.5			
FeO				0.27	1.49
Na_2O				0.88	4.20
K_2O				0.61	4.46
H_2O				trace	trace
Misc.	1.60	1.6			

minute. Active foils were allowed to decay below 15,000 counts per minute, although less than 10,000 counts per minute was generally used. The average of backgrounds taken before and after each set of measurements was used, and counting rates less than twice the average background were discarded.

Fast Neutrons—Aluminum and Iron Foils

As calculated from Bethe's penetrability function,[3] the (n, p) reaction reaches half its maximum cross section for Al and Fe at 4.5 and 7.6 Mev, respectively. Al also has an (n, α) reaction reaching half the maximum cross section at 8.3 Mev. A single set of Al foils gives results for both energy intervals. Reynolds 24-mil (165 mg/cm²) Al sheets, with a purity of 99.8 percent, were die-cut into disks $1\frac{3}{16}''$ in diameter. Fe foils were die-cut from Armco Magnetic Ingot Iron into disks of the same diameter from sheets hand-rolled to 10 mil (200 mg/cm²). After activation, the Al foils exhibited only the 10.2 min. half-life from $Al^{27}(n, p)Mg^{27}$ and the 14.8 hr. half-life[4] from $Al^{27}(n, \alpha)Na^{24}$, indicating that any impurities were negligible. The 2.59 hr. activity[5] from $Fe^{56}(n, p)Mn^{56}$ was observed but in addition there was an apparent half-life of about 28 hr. A correction was made for the 28 hr. period.

To reduce the effect of (n, γ) reactions, the Al and Fe foils were encased in 1 mm thick Cd holders. In the Fe foils, no (n, γ) activation was detectable. By waiting 20 min. before counting, the 2.4 min. activity in Al became negligible.

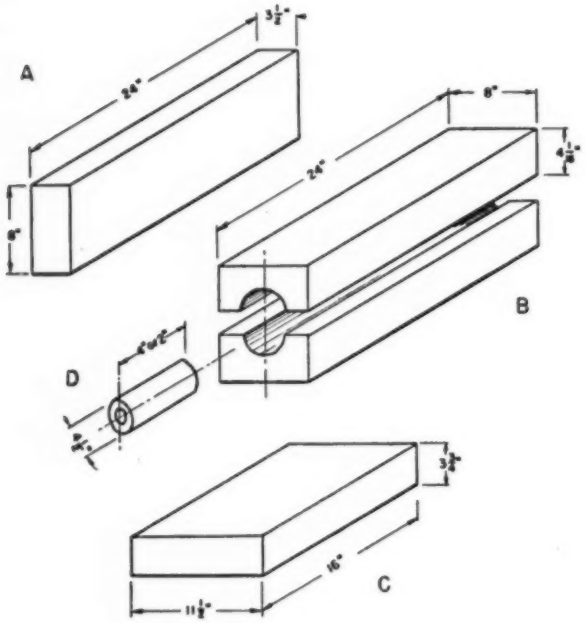


FIG. 3. Special blocks used to form measuring hole and to hold foils and film.

[3] H. A. Bethe, Rev. Mod. Phys. **9**, 69 (1937).
[4] K. Way, *et al.*, Nucleonics **2** (No. 5), 104 (1948).
[5] Segrè Chart, revised April 1948 (AECD-2111).

TABLE II. Average composition of concrete.

H	0.85
O	50.64
Si	30.49
Al	4.44
Ca	8.03
Fe	1.19
Mg	0.23
Na	1.66
K	1.87
S	0.12

Thermal and Resonance Neutrons—Indium Foils

Indium foils were used for measuring thermal and resonance neutron distributions by the Cd-difference method. Foils 1 cm in diameter were cut from 99.97 percent in sheets hand-rolled to a thickness of 3 mil (56 mg/cm²). Activation indicated impurities were negligible. Because the In foils were bombarded only short lengths of time, the 4.5 hr. activity was relatively weak, and correction was made only when it was significantly above background. In foils were counted front and back after bombardment, since the sum of the front and back activity is proportional to the flux for thermal and 1.44-ev resonance neutrons.

Indium Foil Calibration and Corrections

Attempts were made to calibrate the indium foils with the cyclotron as a source, but, owing to nonhomogeneity of the neutron distribution in the cyclotron vault, this was unsuccessful. Instead, a 1-gram Ra-Be source in a 4-foot cubical water tank was used. In addition to corrections for variations in individual foils, two other factors were considered before the foil activities could be interpreted in terms of absolute neutron flux. These factors were (1) the attenuation of 1.44-ev resonance neutrons by the 1-mm Cd holder used for resonance measurements and (2) the distortion of the thermal neutron flux by the detector. Both factors are thoroughly discussed by Bothe[6] and his results are adapted to this work. Bothe develops an expression for the absorption of a neutron striking a detector in any direction:

$$\alpha = 1 - e^{-\mu\delta}(1 - \mu\delta) + \mu^2\delta^2 E_i(-\mu\delta)$$

where α = average absorption probability for a neutron striking in any direction, δ = linear thickness of absorbing material, $\mu = N\sigma$, and E_i = exponential integral.

With this absorption factor, the correction factor to be applied to all resonance runs (Cd holders) becomes:

$$f_a = 1/(1 - \alpha).$$

Inserting the appropriate values for Cd, $\delta = 0.1$ cm $N = 0.0463 \times 10^{24}$ nuclei/cc,[7] and $\sigma_{1.44} = 11.5$ barns,[8] we obtain $f_a = 1.11$.

[6] W. Bothe, Zeits. f. Physik **120**, 437–449 (1943).
[7] H. R. Kroeger, Nucleonics **5** (No. 4), 51–54 (1949).
[8] Goldsmith, Ibser, and Feld, *The Science and Engineering of Nuclear Power*, Vol. I, p. 456 (1947).

To correct for distortion, the depression factor for foils which are small compared with the scattering path length in the medium was taken from Bothe's work using λ_{tr} instead of λ_s:[9]

$$f=\frac{1}{1+0.23\alpha R_K/\lambda_{tr}}$$

where f=depression factor, α=absorption factor, previously described, R_K=radius of foil, λ_{tr}=transport mean free path.

Inserting appropriate values for In, $\alpha=0.100(\sigma_{th}=190$ barns), $R_K=0.5$ cm, $\lambda_{tr}=1.95$ cm, we obtain $f=1/1.00590$, or the correction factor equals 1.00590.

After applying these factors to the resonance and Cd difference results, the corrected, saturated initial activities were converted to absolute neutron flux.

Gamma-Ray Measurements

Measurements of total gamma-radiation proved considerably more difficult than neutron measurements. X-ray films (duPont 552) were used. This film is paper-wrapped in dental-size packets, each containing two films, one sensitive and one insensitive, such that 0.02 to 20 roentgens can be measured. The exposed films are examined with an Ansco-Sweet Densitometer and the density is converted to dosage by means of calibration charts. These charts are based on a 1-mm layer of Cd surrounding the film.

The film has two major undesirable characteristics. The bare film is severely blackened by beta-rays and electrons. Furthermore, the blackening is constant for gamma-radiation above 300 kev. Below 300 kev, the blackening per roentgen rises sharply. When only alpha-, beta- and gamma-radiations are present, 1 mm of Cd is sufficient to attenuate the charged radiations and has been shown to be adequate for reducing the blackening from the soft gamma-radiation. When there is a high thermal neutron flux, as in most of the present measurements, the secondary gammas from radiative capture of neutrons in Cd produce serious extraneous blackening. Kikuchi, Aoki and Husimi[10] found that of 32 elements studied, Cd is outstanding in its production of capture gammas.

Summary of Runs

In order subsequently to refer to runs by number those providing useful data are listed in Table III.

RESULTS AND DISCUSSION

Fast Neutron Distribution

For the fast neutrons we assume a point source at the cyclotron target and an exponential attenuation through the wall, i.e., a $(1/r^2)\exp(-r/L_r)$ distribution,[11] where r is the distance of the foil from the target and L_r is the relaxation length at large distances.

Figure 4 shows one set of measurements of fast neutron attenuation through the concrete wall for Al and Fe detectors. We would expect the relaxation lengths for these curves to be the same for the three activities observed since the energies involved are nearly the same. At large distances in the wall, it would appear that the fast, resonance, and thermal neutrons should have the same relaxation length, since the higher energy feeds into the lower energies and the diffusion length of low energy neutrons is quite small. Table IV shows the results.

Values for both thresholds of the Al foils agree and are considered the most acceptable for the threshold range covered. The attenuation factor corresponding to a relaxation length of 14.4 cm is 8.3 per foot.

Dementi and Timoshuk's[12] cross section for the Al(n, p) reaction is 15 millibarns and the extrapolated short-lived activity at the wall's outer surface is 5.75 c.p.m. for Run 14. This run is used because its average beam current was the strongest of the Al runs, 164 μamp., and its relaxation length the same as the average, 14.4 cm.

The flux is estimated from the following relation:

$$(nv)_f=A/(\sigma_p NV),$$

where $(nv)_f$=flux in neutrons/cm^2/sec., A=absolute foil activity in disintegrations per second, σ_p=reaction cross section in cm^2, N=nuclei per foil volume, and V=foil volume.

This gives a flux of 243 neutrons/cm^2/sec. above the Al(n, p) threshold and gives an interesting comparison with the tolerance flux of 190 n/cm^2/sec. for 4.5-Mev neutrons determined from:[13]

$$(nv)_f=1.45\times10^{-21}/\sigma_H E,$$

TABLE III. Summary of runs.

Run no.	Detector	Type run	Length	Av. beam current (μ amp.)	Target	Remarks
1	Al	*A*	2 h	60	Be	
4	Al	*B*	20 m	96	Be	
9	Fe	*B*	11½ h	155	Cu	
13	Al	*B*	34 m	150	Be	
14	Al	*A*	30 m	164	Be	
15	In	*A*	10 m	165	Be	Thermal+Resonance (Paper holders)
16	In	*A*	10 m	170	Be	Resonance (Cd holders)
17	Film	*A*	5 m	168	Be	
18	Film	*A*	10 m	170	Be	
19	Film	*B*	5 m	172	Be	
20	Film	*B*	10 m	168	Be	
21	Fe	*B*	6½ h	165	Be	
22	Fe	*A*	7h-55 m	164	Be	
23	In	*B*	10 m	170	Be	Thermal+Resonance
24	In	*B*	10 m	162	Be	Resonance

[9] This significant correction was suggested by C. W. Tittle.

[10] Kikuchi, Aoki, and Husimi, Proc. Phys. Math. Soc. Japan 18, 115 (1936).

[11] R. F. Christy, MDDC-1175, p. 127.

[12] V. S. Dementi and D. V. Timoshuk, Comptes Rendus Acad. Sci. U.R.S.S. **27**, 926 (1940).

[13] M. L. Goldberger, *The Science and Engineering of Nuclear Power* Vol. II, Chapter 12 (1949).

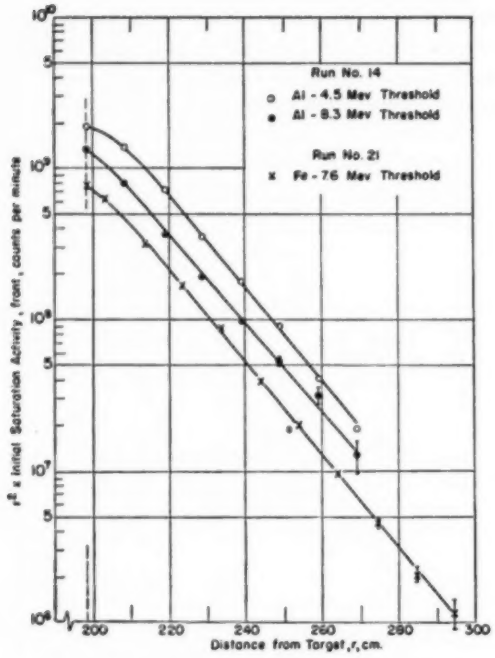


FIG. 4. Fast neutron distributions in concrete.

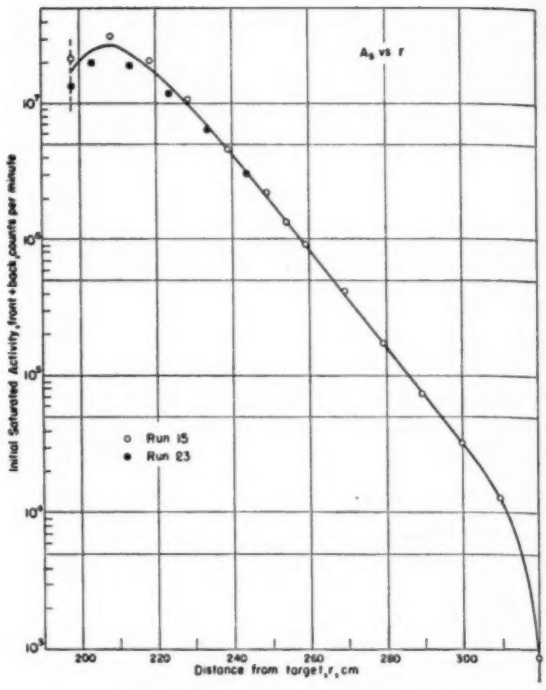


FIG. 5. Thermal plus 1.44 ev resonance neutron distributions in concrete.

where σ_H = scattering cross section for hydrogen and E = neutron energy in Mev. Our approximate value can also be compared with the tolerance flux of 85 n/cm²/sec. for 2–20 Mev neutrons reported by Mitchell.[14]

Our estimate is probably low, due to the decrease in reaction cross section as the Al(n, α) reaction begins to compete.

Thermal and Resonance Neutron Distributions

Since the cyclotron is non-reproducible, it was necessary to establish a relation between the resonance and thermal resonance distributions in order to provide a basis for normalization.

Table III shows that 15 and 23 were thermal+resonance runs and 16 and 24 were resonance runs. A foil was placed 5½" to the right of the measuring hole axis on the inner face of the wall; on Runs 15 and 23, the foil was Cd covered, on Runs 16 and 24, paper covered. Between each of these specially placed foils and the regular foil nearest the cyclotron, thermal+resonance/resonance ratios were established. Since only one foil was covered with Cd on Run 23, with the depressive effect consequently less, the thermal+resonance/resonance ratio of this run (2.594) was selected to be applied to the mean values of the thermal+resonance and resonance runs in order to relate them.

TABLE IV. Relaxation lengths for Al and Fe threshold detector runs, from r^2A_s vs. r plots (in cm).*

Al (4.5 Mev)	Al (8.3 Mev)	Fe (7.6 Mev)
14.1 (1)	14.1 (1)	14.7 (9)
13.7 (4)	14.1 (4)	14.0 (21)
15.2 (13)	14.7 (13)	13.4 (22)
14.4 (14)	14.6 (14)	
Av. 14.4	14.4	14.0

* Numbers in parentheses indicate the number of the run.

[14] J. S. Mitchell, Brit. J. Radiology **20**, 177 (1947).

Figure 5 shows a plot of the results of Runs 15 and 23, with 15 corrected to an average beam current of 170 μamp. by assuming the neutron yield to be linearly proportional to the beam current for small increments of current, together with the mean curve for the two runs. Figure 6 shows a plot of the results of Runs 16 and 24, with 24 corrected to 170 μamp. beam current and the resulting mean curve. The thermal+resonance/resonance ratio of 2.594 was applied to the initial point of the mean resonance curve of Fig. 6 to determine the starting point of the corresponding thermal+resonance curve. A point-by-point subtraction was carried out next and the foil depression correction applied to this result, giving the thermal distribution shown in Fig. 6.

Mean Square Distances and Diffusion Length

One of the first results obtainable from the thermal resonance data is the calculation of the mean square distance for thermal and resonance neutrons. Before reasonable values can be determined, however, an initial assumption is necessary for which no justification can be found in theory. The primary source of neutrons is naturally the cyclotron target, and at first glance it seems correct to consider a distribution based on the actual distance of each foil from the target, following

the usual $(1/r^2)\exp(-r/L_r)$ attenuation, where r is the distance from the target and L_r the relaxation length. Figure 7 shows such a plot. Amaldi and Fermi's equation can then be used to calculate mean square distance:[15]

$$\bar{r}^2=\frac{\int_0^\infty r^4 A dr}{\int_0^\infty r^2 A dr},$$

where r is the distance of the foil from the target and A is the foil activity as a function of distance.

Values determined by this method for the mean square distance for thermal and resonance neutrons are absurdly high, 5×10^4 cm^2, and have no significance, being obviously overweighted by the effect of the 200-cm distance in air between the target and the front face of the wall. By assuming that the thermal and resonance neutron source is located at the inner face of the wall, but continuing to treat it as a point source (as plotted in Fig. 8), reasonable values are obtained:

$$\bar{r}_{th}^2=2148 \text{ cm}^2 \text{ and } \bar{r}_{res}^2=1818 \text{ cm}^2.$$

The root mean square distance for the resonance neutrons, $\bar{r}_{res}=42.7$ cm, compares closely with Jensen and Ritter's value of 44.9 cm,[16] which is corrected for boundary disturbance and also is based on the use of a Ra-Be neutron source. Our result suggests a lower mean energy for the neutron spectrum of the cyclotron than for a Ra-Be source.

Continuing, we estimate the diffusion length L from the relation:[17]

$$\bar{r}^2_{thermal}=\bar{r}^2_{thermal\ source}+6L^2.$$

Although $\bar{r}^2_{thermal\ source}$ is required by the equation, we can accept the value of $\bar{r}^2_{res}$ as a close approximation. From this we obtain a diffusion length of 7.4 cm.

We can compare this with the calculated results: $L^2=\lambda_{tr}\lambda_a/3$, where λ_{tr} and λ_a are the transport and absorption mean free paths, respectively. Table V lists the cross sections used in computing λ_a, λ_s and λ_{tr}. The Mg and S in the concrete were neglected.

Table VI lists the calculated values of absorption, scattering and transport mean free paths, and the diffusion length. The value of the latter, 7.7 cm, is in good agreement with the experimental observed value of 7.4 cm given in Table VI.

Relaxation Lengths for Thermal and Resonance Neutrons

Assuming a thermal and resonance source at the inner face of the wall, we obtain relaxation lengths of

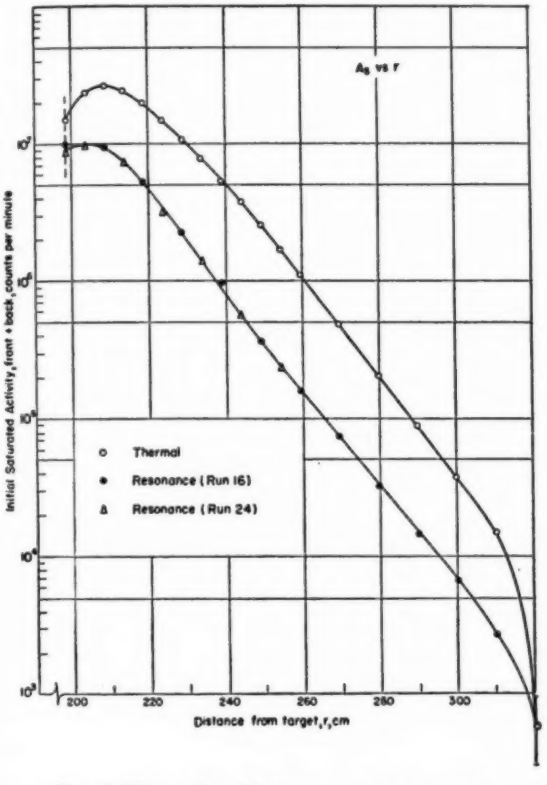


FIG. 6. Thermal and 1.44 ev resonance neutron distributions in concrete.

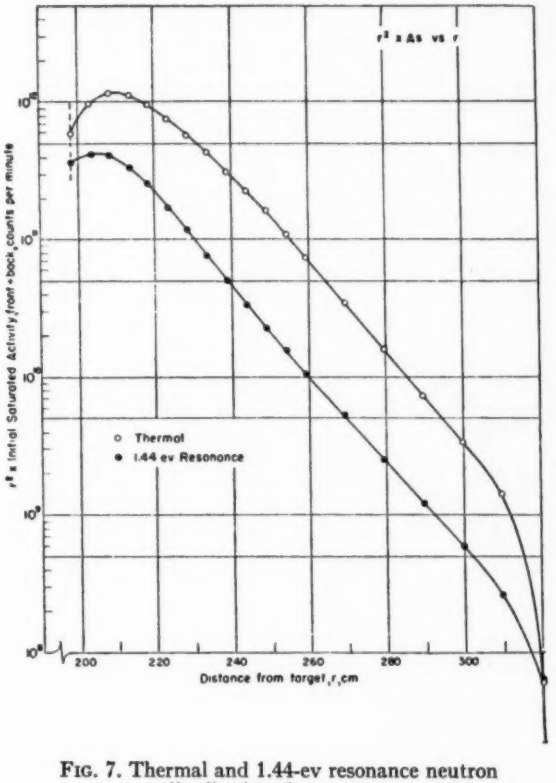


FIG. 7. Thermal and 1.44-ev resonance neutron distributions in concrete.

[15] E. Amaldi and E. Fermi, Phys. Rev. **50**, 899–928 (1936).

[16] P. Jensen and O. Ritter, Zeits. f. Naturforschg. **2a**, 376–384 (1947).

[17] A. M. Munn and B. Pontecorvo, Can. J. Research **25A**, 157 (1947).

$L_{\tau(th)} = 16.9$ cm and $L_{\tau(res)} = 18.0$ cm. Jensen and Ritter[18] embedded their source in the center of a concrete block and reported relaxation lengths of 20.4 cm and 24 cm for thermal and 1.44-ev resonance neutrons, respectively. It should be noted that they used a half-life for In of 57 minutes, which would account for some of the discrepancy between our measurements and theirs. Their experimental arrangement was close to ideal, making it unnecessary to adopt the assumption of the source at the surface of the wall.

In order to compare our work with recent observations by Gugelot and White,[19] we consider our source as the cyclotron target and the thermal and resonance relaxation lengths thus obtained are 13.1 and 14.0 cm, respectively. Gugelot and White, using Ag detectors, report a half-value thickness of 9.5 cm for ordinary concrete with 8.5 percent water content. This corresponds to a relaxation length of 13.7 cm.

The relaxation length for thermal neutrons, based on the cyclotron target as the source, is equivalent to an attenuation factor of 10 per foot for slow neutrons, distinctly larger than the 8.3 per foot obtained for fast neutrons. Since the former is based on the observed thermal distribution, which is considerably more uncertain, a value of 9 per foot is considered a reasonable average for future design purposes.

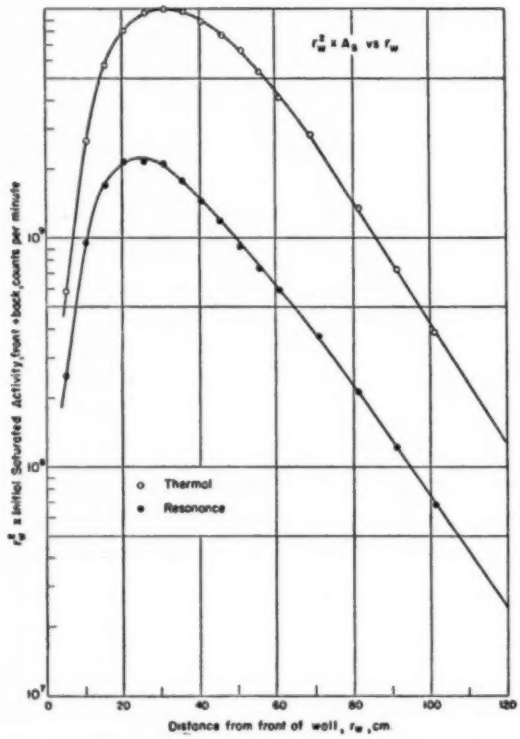


FIG. 8. Thermal and 1.44-ev resonance neutron distributions in concrete, with respect to front of wall.

Absolute Thermal and Resonance Neutron Fluxes

In the same four-foot water tank used for In foil calibration, Dacey and Paine[20] have accurately established the thermal and resonance flux at a point 14 cm from a 1-gram Ra-Be source. From the activities of In foils at this point, we obtain the following ratios for converting In foil activity from a cyclotron bombardment to neutron flux:

$$\frac{(nv)_{th}}{(A_s)_{th}} = \frac{8613}{6082} = 1.416,$$

$$\frac{(nv)_{res}}{(A_s)_{res}} = \frac{124.2}{483.5} = 0.2569.$$

These are then applied to the mean values for the thermal and resonance distributions, respectively, to obtain the thermal flux in neutrons/cm²/sec. and the resonance flux in neutrons/cm²/sec. per unit logarithmic energy interval.

It is interesting to compare the thermal flux at the outer surface of the wall with values calculated by Wende[21] and by Mitchell[22] as biologically equivalent to the *former* tolerance dosage of 0.1 r/8 hour day.[23] Considering only the effect on body tissue of gamma-rays from the $H'(n, \gamma)$ reaction, Wende estimates the tolerance flux to be 15,000 n/cm²/sec. for thermal neutrons. Mitchell, in a more realistic calculation, includes the

TABLE V. Thermal cross sections for nuclei in concrete.*

Nucleus	(Nuclei/cc) $\times 10^{24}$	σ_a	σ_s (in barns)	σ_{tr}
H	0.0112	0.32	60.4	26.6
O	0.0419	0.001	4.1	3.9
Si	0.0144	0.2	2.3	2.2
Al	0.0022	0.22	1.4	1.4
Ca	0.0027	0.5	3.5	3.4
Fe	0.00028	2.5	11	11
Na	0.00096	0.45	4	4
K	0.00063	2.5	1.5	1.5

* σ_s and σ_{tr} for H are from J. E. Dacey and R. W. Paine, Jr., S.M. Thesis, M.I.T. (1949). All other cross sections from H. R. Kroeger, Nucleonics 5 (No. 4), 51–54 (1949).

TABLE VI. Absorption, scattering and transport mean free paths and diffusion lengths (in cm).

λ_a	90.9
λ_s	1.11
λ_{tr}	1.95
L (calculated)	7.7
L (experimental)	7.4

[18] P. Jensen and O. Ritter, Zeits. f. Naturforschung **2a**, 376–384 (1947).

[19] P. C. Gugelot and M. G. White, Phys. Rev. **74**, 1215 (1948). Subsequent to the submission of the present manuscript, an article by these authors has appeared giving more complete data, J. App. Phys. **21**, 369 (1950).

[20] J. E. Dacey and R. W. Paine, Jr., S.M. Thesis, M.I.T. (1949).

[21] C. W. J. Wende, *The Computation of Radiation Hazards* Report No. M-1324 (TNX-7) (11 January 1944).

[22] J. S. Mitchell, Brit. J. Radiology **20**, 79 (1947).

[23] R. E. Lapp and H. L. Andrews, *Nuclear Radiation Physics* (Prentice-Hall, Inc., New York, 1948), Chapter 18.

effects of the $N^{14}(n, \gamma)$ and $N^{14}(n, p)$ reactions as well as the gammas from neutron capture by hydrogen and estimates the tolerance flux to be 12,500 n/cm²/sec. Our data indicate the thermal flux at the wall's outer surface to be 830±540 n/cm²/sec. It is evident, therefore, that for 170 μamp. average beam current and a Be target, the cyclotron shield has reduced the thermal flux more than an order of magnitude below either of these estimated tolerance levels and well below the recent more conservative level of 1500 n/cm²/sec.[24]

Gamma-Radiation

Figure 9 is typical of radiation measurements using bare film. The plot is in terms of $r^2 \times I$ as a function of r, where r is the distance in cm from the cyclotron target to the film and I the apparent ionization intensity in roentgens per hour. The values for I are determined as previously described. These intensities are somewhat uncertain because of the possible blackening caused by low energy gamma-rays (<300 kev) and by capture gammas from the cadmium cover.

Two prominent features of all the curves obtained are the increase in apparent ionization intensity in the first 10 cm of the wall because of secondary and capture gammas from neutron interaction within the wall and also the tendency toward linearity beyond 40–45 cm of wall distance.

At large distances in the shield three processes combine to produce the total gamma-radiation: (1) attenuation of primary gammas from the target (2–8 Mev); (2) capture gammas from thermal neutron capture by hydrogen (2.18 Mev); (3) secondary gammas from radioactive products of fast neutron reactions (0.5–2.8 Mev).[25] The relative weights of these effects cannot be determined, although (1) and (2) appear more important than (3).

From the curves, an average absorption coefficient of 0.058 cm^{-1} is obtained. Hirschfelder and Adams[26] have reported mass absorption coefficients for monoenergetic gamma-rays in concrete. Converting their results for 1- and 3-Mev gamma-rays into linear absorption coefficients, we obtain values of 0.0856 and 0.0517 cm^{-1}. Assuming the ionization intensity is directly proportional to the flux intensity over the energy range with which we are concerned, the absorption coefficient from our data suggests a "mean" gamma-ray energy of about 2.7 Mev. Gugelot and White,[27] using 18-Mev protons on Be, have reported a half-value thickness of 13.3 cm for gamma-rays in concrete of slightly higher density (2.35 g/cc) than used here. This corresponds to an absorption coefficient of 0.052 cm^{-1} and, according to Hirschfelder and Adams,[26] to a "mean" energy of about 3 Mev. The attenuation factor for the gamma-radiation in concrete, measured in the linear portion of the curves, is 5.9 per foot.

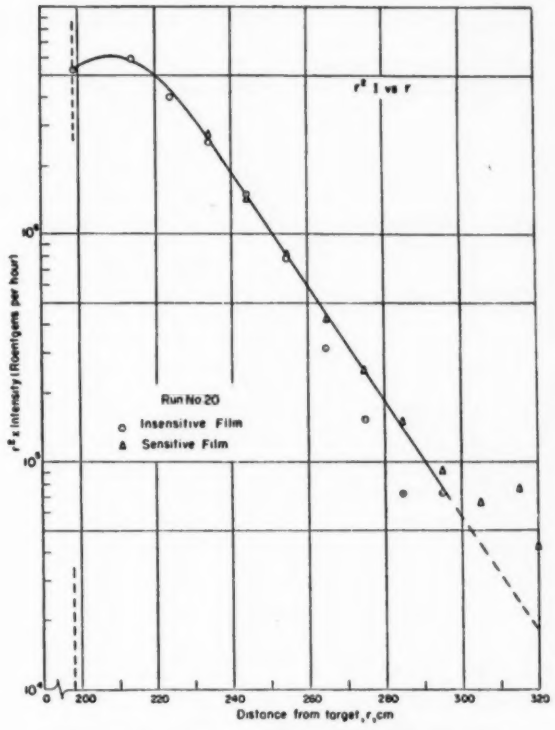


FIG. 9. Apparent ionization intensity distribution through concrete.

Average gamma-radiation from the cyclotron is probably about 5 Mev. The capture and secondary gammas, with lower average energies, appear sufficiently abundant to result in a lower mean energy for the total gamma-radiation in the wall. This seems to agree with Gugelot and White's[27] results.

This effect should be an important consideration in designing future cyclotron shields. With better information, the apparent importance of the capture and secondary effects may be verified. If so, cyclotron shield design should emphasize the importance of lower energy gamma-rays as much as those of higher energies produced at the target.

ACKNOWLEDGMENTS

The authors appreciate discussions of this work with Dr. M. Stanley Livingston and our colleagues in the Nuclear Shielding Group. The facilities of the cyclotron were made available by Dr. R. D. Evans with the assistance of Mr. J. B. Bulkley. The painstaking assistance of Miss Joan Sullivan in the analysis of data is gratefully acknowledged. This work was performed in the Laboratory for Nuclear Science and Engineering and was supported by the Bureau of Ships and the Office of Naval Research.

[24] W. S. Snyder, Nucleonics **6**, 49 (Feb. 1950).
[25] Segrè Chart, revised April 1948, *The Science and Engineering of Nuclear Power*, Vol. II (1948).
[26] J. O. Hirschfelder and E. N. Adams, Phys. Rev. **73**, 863 (1948).
[27] P. C. Gugelot and M. G. White, Phys. Rev. **74**, 1215 (1948).

Storage of Small Signals on a Dielectric Surface

JOHN V. HARRINGTON
Air Force Cambridge Research Laboratories, Cambridge, Massachusetts
(Received May 29, 1950)

A mathematical analysis is presented which is believed to be applicable to a general class of storage tubes where signal storage is accomplished by depositing through secondary emission a charge pattern on a dielectric surface. The assumptions made to linearize and simplify the problem are outlined and plots are given of the predicted output signals for writing, reading, and cancellation operations when the input signal is a step function. Experimental evidence is presented to substantiate the analytical results.

INTRODUCTION

AN increasing amount of material has appeared in recent literature concerning the general class of vacuum tubes called storage tubes, in which signal storage is accomplished by depositing, through secondary emission, a charge pattern on a dielectric surface.[1–5] The material published on these storage tubes thus far has emphasized more or less the physics of the storage process—an emphasis which is to be expected in view of the admittedly complicated process of charge transfer by secondary emission and the difficulty in collecting the secondary current to give a useful output signal. To the engineer using these storage tubes, as opposed to the tube designer, somewhat different characteristics are of primary interest. As in ordinary vacuum tubes, these are external or equivalent characteristics and are defined primarily on the basis of the device's behavior as a circuit element; i.e., on the basis of its effect on the amplitude or shape of a signal passed through it.

The analysis which follows attempts to explain quantitatively some of the external characteristics of the *barrier-grid storage tube*, and a substantial portion of the introductory discussion concerns this particular model. The results which are obtained, however, are believed to be sufficiently general as to apply to the general class of tubes under discussion.

A simplified sketch of the *barrier-grid tube* is shown in Fig. 1. A complete description of this device and its operation appears in the literature.[1] In operation the dielectric surface will be charged, in the absence of a signal, to some potential V_0 at which the effective secondary emission ratio, δ, equals 1 and equilibrium exists. Thereafter when a signal is applied to the back plate, the emission ratio is changed; this results in a charge, different from the equilibrium charge, being deposited on the dielectric at the point of beam impingement. Thus, for a linear beam motion a signal varying in time would be stored as a charge distribution varying in distance along a line of the dielectric surface. In reading, when no signal is applied to the back plate, the beam in sweeping over this same line acts to restore the charge distribution to the equilibrium value, and in so doing the secondary emission current (as observed at the collector) will regenerate essentially a counterpart of the stored signal.

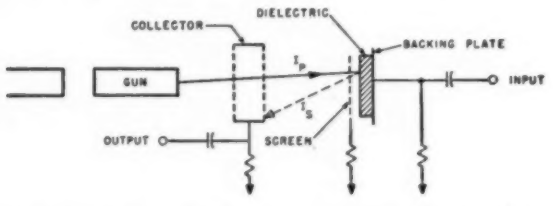


FIG. 1. Barrier grid storage tube, simplified sketch.

BASIS OF ANALYSIS

A rigorous mathematical analysis of a device such as this is extremely difficult if not impossible to make. Consequently certain assumptions and simplifications must be made to bring the problem within the range of available analytical methods. In this case the following assumptions have been made:

(1) The loss of resolution due to the redistribution of secondaries is neglected. The screen action is assumed to reduce such losses to negligible proportions.

(2) Not all of the secondary beam leaving the dielectric gets to the collector. It is assumed, however, that the fraction of the secondaries which are intercepted by the screen and other electrodes is constant; thus the collector or signal current is a "scaled-down" version of the total secondary current.

(3) A linear relationship is assumed between small voltage changes about the second crossover, or equilibrium, point of the secondary emission characteristic, and the corresponding secondary emission ratio (Fig. 2).

(4) The current in the electron beam is assumed to follow an even distribution curve in the direction of motion of the beam and

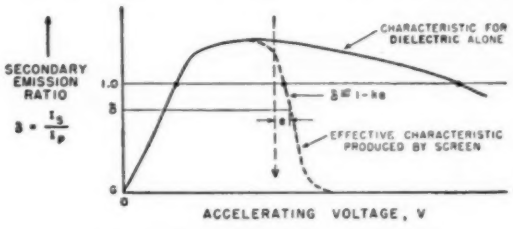


FIG. 2. Secondary emission characteristics, with and without screen.

[1] Jensen, Smith, Mesner, and Flory, RCA Review (March, 1948).
[2] R. A. McConnell, Proc. I.R.E. (November, 1947).
[3] A. V. Haeff, Electronics (September, 1947).
[4] H. Klemperer, J. T. de Bettencourt, Electronics (August, 1948).
[5] V. K. Zworykin, G. A. Morton, Television (John Wiley and Sons, Inc., New York, 1939).

to be uniformly distributed in the transverse direction. The effect of this assumption is to reduce what would be a three-dimensional problem to one of two dimensions—distance and time. This may be done with small sacrifice of generality, since it is primarily the lateral and not the transverse distribution of the beam current which influences the signal response. For purposes of presenting the results of this analysis graphically a normal beam distribution is assumed.

(5) Effects such as unevenness or non-uniformity of the dielectric, secondary emission from the screen material, etc., which contribute primarily to the background noise level of the tube, are ignored.

The foregoing assumptions allow us to reduce the problem to that of evaluating the current-voltage relationships in the equivalent network shown in Fig. 3. $e_0(t)$ is the signal voltage on the back plate applied at time $t=0$; C_x is the capacitance of the dielectric per unit area; and δ is the effective secondary emission ratio and is given by $\delta=1-ke$ for small values of e.

The quantity I_x is the primary beam-current density at point x; if we assume that the beam is scanning with constant lateral velocity v, then I_x as a function of x

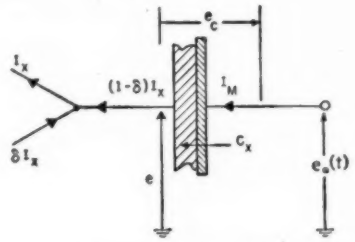


FIG. 3. Storage-tube input circuit.

and t is given by:

$$I_x=\frac{I_b}{w}F\left(\frac{x-x_0-vt}{w}\right)=\frac{I_b}{w}F(x'-t')$$

where $F(x')$ is an even function of x' and meets the conditions:

$$\int_{-\infty}^{\infty}F(x')dx'=1 \text{ so that } \int_{\infty}^{\infty}Idx=I_b$$

and for $|x'|\gg 1$, $F(x')\to 0$.

I_b is the total beam current per unit line width and w is a measure of a half beam width. The primed variables are normalized or dimensionless units of distance and time are given by

$$x'=(x-x_0)/w, \quad t'=vt/w.$$

For the circuit of Fig. 4 we may write the following equations

$$e=e_0(t)-e_c, \tag{1}$$

$$C_x(de_c/dt)=(1-\delta)I_x. \tag{2}$$

Then the basic differential equation obtained by combining the foregoing relationships is

$$de_c/dt'+\gamma F(x'-t')e_c=\gamma F(x'-t')e_0(t)$$

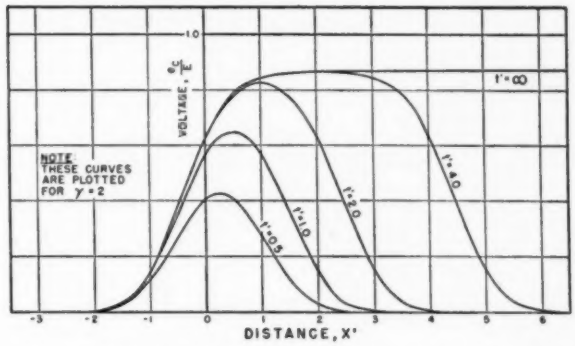


FIG. 4. Build-up of voltage distribution along scan line during storage of step voltage.

where

$$\gamma=[kI_b/vC_x]. \tag{3}$$

Since the signals to be stored in any practical case are most always pulsed signals, it is useful to carry out the analysis on the assumption that $e_0(t)$ is a step function of magnitude E applied at $t=0$. From these results one can easily deduce the tube performance for pulsed signals. Under these conditions (3) has the general solution

$$e_c(x',t')=E+C_1(x')\epsilon^{\gamma[B(x'-t')-B(x')]}, \tag{4}$$

where

$$B(x')=\int_0^{x'}F(x')dx'$$

and $C_1(x')$ is a constant of integration (the integration having been performed with respect to t') to be determined by the initial conditions.

The varying component of the collector current which generates the output signal will be the negative of the dielectric charging current I_M; it is calculated as follows:

$$I_M=\int_{-\infty}^{\infty}(1-\delta)I_xdx=\int_{-\infty}^{\infty}C_x\frac{de_c}{dt}dx=vC_x\int_{-\infty}^{\infty}\frac{de_c}{dt}dx. \tag{5}$$

In setting the limits for which this integration is to be performed, the implication is made that the length of a scan line is infinite. This is a valid assumption so long as the scan line is very long compared to one beam width.

ANALYSIS

It is both convenient and useful to obtain solutions for the charging current and for the voltage distribution along a scan line for three general cases. The first case will correspond to the so-called writing operation in which a voltage step is to be stored with no signal initially in storage; the second case will correspond to a multiple writing or cancellation operation in which a second-step voltage is to be stored over one initially in storage; the third case corresponds to a reading operation in which no external signal is applied but a stored signal is present.

Case I. Writing Operation

For this case the initial conditions are at $t=0$, $e_c=0$, $e_0(t)=E$, when these are used to evaluate $C_1(x)$, the desired expression for e_c is:

$$e_c(x', t')=E[1-\epsilon^{-\gamma[B(x')-B(x'-t')]}]. \quad (6)$$

It is interesting to plot this (Fig. 4) for an assumed[6] normal beam current distribution in order to observe the changing charge pattern laid down by the electron beam as the applied step voltage is stored. In interpreting these curves it is helpful to consider t' as a distance giving the center of the beam. For very large values of t', the voltage distribution especially in the vicinity of $x'=0$ becomes essentially independent of time and is given by

$$\operatorname{Lim}_{t'\to\infty} e_{c1}(x', t')=e_{c1}(x', \infty)=E_1[1-\epsilon^{-\gamma[\frac{1}{2}+B(x')]}]. \quad (7)$$

This expression represents the "stored" signal.

Case II. Multiple Writing Operation

By using (7) as the initial voltage distribution for the second writing operation in which a step of magnitude E_2 is to be stored, it is found that

$$e_{c2}(x', t')=E_1[1-\epsilon^{-\gamma[\frac{1}{2}+B(x')]}]\epsilon^{-\gamma[\frac{1}{2}-B(x'-t')]}; \quad t\leq 0 \quad (8)$$

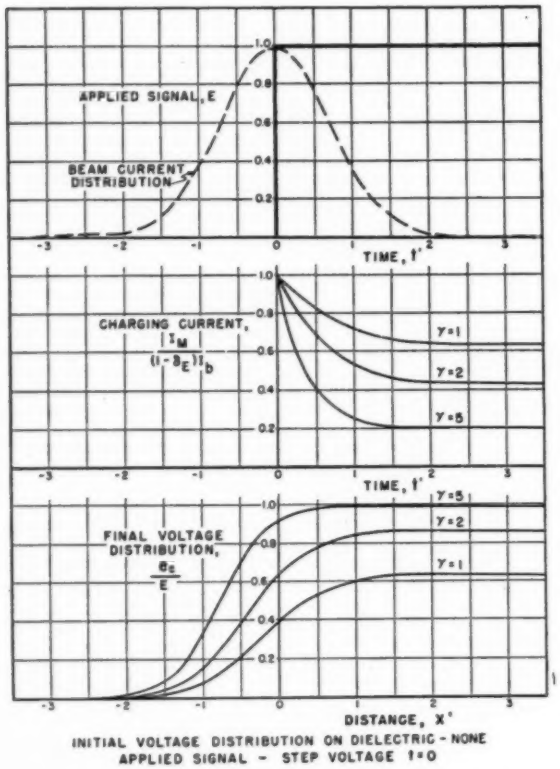


FIG. 5. Signal storage on a dielectric surface, Case I.

[6] That is, in the curves, $F(x')=[I_b/(\pi)^{\frac{1}{2}}w]\epsilon^{-x'^2}$, and $B(x')=\frac{1}{2}\operatorname{erf}(x')$.

$$e_{c2}(x', t')=E_2[1-\epsilon^{-\gamma[B(x'-t')-B(x')]}]+E_1\epsilon^{-\gamma}[\epsilon^{\gamma[\frac{1}{2}+B(x')]}-1]\times\epsilon^{\gamma[B(x'-t')-B(x')]}; \quad t\geq 0. \quad (9)$$

The final "stored" signal obtained by letting $t'\to\infty$, is then

$$e_{c2}(x', \infty)=(E_2+\epsilon^{-\gamma}E_1)[1-\epsilon^{-\gamma(\frac{1}{2}+B(x'))}]. \quad (10)$$

From (8), (9) and (5) one finds that the charging current is given by

$$I_M=kE_1I_b\cdot\epsilon^{-\gamma}\left[I(\gamma, t')-\left(\frac{\epsilon^{\gamma}-1}{\gamma}\right)\right]; \quad t\leq 0 \quad (11)$$

$$I_M=kI_b\cdot\left[(E_2+E_1\epsilon^{-\gamma})I(\gamma, t')-E_1\left(\frac{1-\epsilon^{-\gamma}}{\gamma}\right)\right]; \quad t\geq 0 \quad (12)$$

where the function $I(\gamma, t')$ is given by:

$$I(\gamma, t')=\int_{-\infty}^{\infty}\epsilon^{\gamma[B(x')-B(x'-t')]}F(x')dx'. \quad (13)$$

Equations (11) and (12) are rather general expressions for the dielectric charging current and as special cases of these one may obtain by setting $E_1=0$, $E_2=E_1=E$, or $E_2=0$, the expression for I_m for cases I, II, or III,

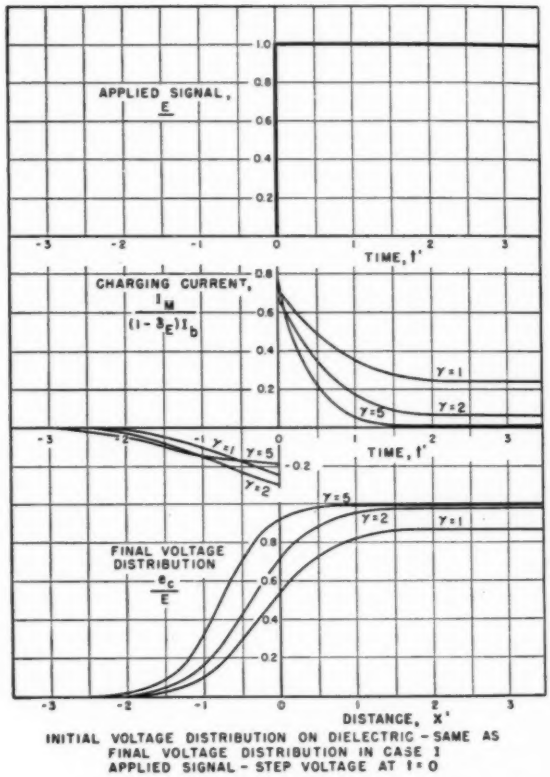


FIG. 6. Signal storage on a dielectric surface, Case II.

respectively. It is interesting to note that while a simple relation giving $I(\gamma, t)$ is in general impossible to obtain because of the difficulty encountered in integrating (7), its values at $t=0$, and $t=\pm\infty$ can be calculated.

$$I(\gamma, 0)=1$$
$$I(\gamma, \infty)=(1-\epsilon^{-\gamma})/\gamma$$
$$I(\gamma, -\infty)=(\epsilon^{\gamma}-1)/\gamma. \quad (14)$$

Thus we may conclude that the particular beam distribution, i.e., the particular $F(x')$; only influences the shape of the response within a few beam widths travel from the point of application of the signal.

Even within this region the $I(\gamma, t')$ functions for different assumed beam current distributions are very much alike. Beyond this region, the response (a sort of steady state response) is independent of $F(x')$ and depends principally on γ. This is to be expected since for a flat or uniform response only the total beam current and not its distribution should be significant.

The charging current and final voltage distribution for selected values of γ and an assumed normal beam current distribution for cases I and II are plotted[7] in Figs. 5 and 6. The typical "spike" for the writing signal is indicated, and it is seen that the effect of the finite width of the electron beam is to slope the leading edge of the stored signal such that two modified distance units (roughly, one beam width) are required to go from initial to final values.

Case III. Reading Operation

The expressions for charge distribution and charging current for the reading operation may be easily obtained from (10)–(12), by setting $E_1=E$ and $E_2=0$. A comparison of (10) and (7) indicates that for the reading operation the initial and final voltage distributions would be identical except for a scale factor $\epsilon^{-\gamma}$. Thus the reading operation is also an erasing operation in which the stored signal is reduced by $\epsilon^{-\gamma}$ or $8.68\gamma db$.

Typical curves for the Case III charging current are plotted in Fig. 7 along with the corresponding voltage distribution curves for the reading operation. It is seen that the output current wave rise time has been limited to about 3-to-4 modified time units or to the time it takes the beam to sweep through 1.5 to 2 beam widths.

EXPERIMENTAL RESULTS

As intimated previously the purpose of this section is to present some experimental evidence obtained with the STE barrier-grid storage tube which will support the preceding analysis. To obtain this evidence it is first necessary to choose a write-read schedule which on a continuous or repetitive basis will yield signal shapes corresponding to those predicted for Cases I, II and III. It turns out that about the simplest schedule to do this is two writing followed by two reading operations, ad infinitum (Fig. 8). Because of the repeated application of the signal and the improbability of complete erasure, the initial charge distribution prior to the application of the first writing pulse will not be zero as assumed in Case I but will have some finite value $AE(x)$. It is desirable to calculate the magnitude of A as a function of γ so that our results may be properly interpreted. Thus, from the preceding analysis,

if before *first write*, stored signal is $A\cdot E(x')$ (15)

then before *second write*, stored signal is

$$(1+\epsilon^{-\gamma}A)\cdot E(x') \quad (16)$$

and before *first read*, stored signal is

$$[1+\epsilon^{-\gamma}(1+\epsilon^{-\gamma}A)]\cdot E(x') \quad (17)$$

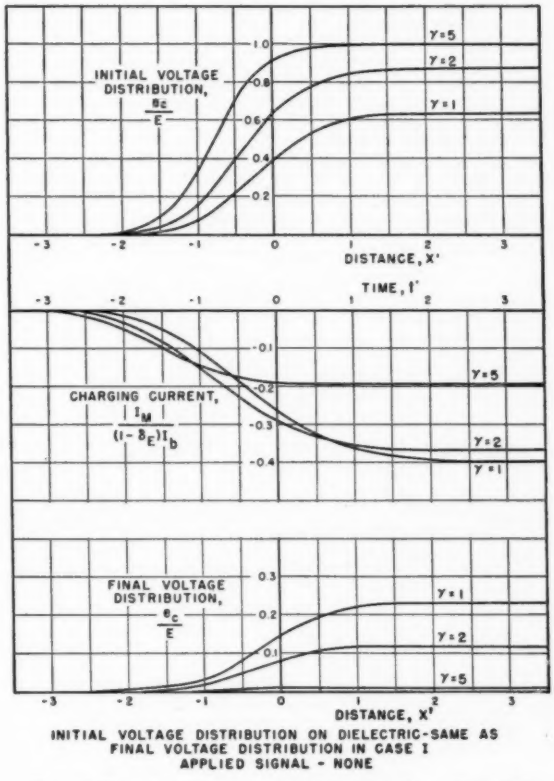


FIG. 7. Signal storage on a dielectric surface, Case III.

and before *second read*, stored signal is

$$\epsilon^{-\gamma}[1+\epsilon^{-\gamma}(1+\epsilon^{-\gamma}A)]\cdot E(x') \quad (18)$$

and after *second read*, stored signal is

$$\epsilon^{-2\gamma}[1+\epsilon^{-\gamma}(1+\epsilon^{-\gamma}A)]\cdot E(x'). \quad (19)$$

Then for a sustained sequence of such operations equilibrium will be reached when (15)=(19), or when

$$A=\epsilon^{-2\gamma}\frac{(1+\epsilon^{-\gamma})}{(1-\epsilon^{-4\gamma})}. \quad (20)$$

Using this to determine the initial values of the charge distribution, one can utilize the results of the previous

[7] For selected values of γ and for an assumed normal beam current distribution, numerical values of $I(\gamma, t')$ in the range $-4<t'<4$ have been calculated at the M.I.T. Center of Analysis, Cambridge, Massachusetts.

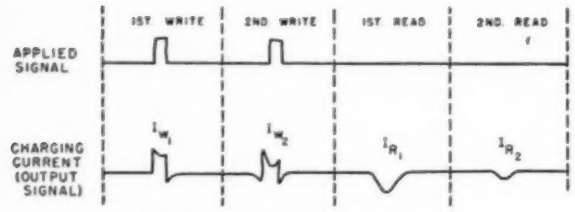


FIG. 8. Two-write two-read schedule.

section, Eq. (12), to find the steady-state responses of the write-read signals of Fig. 8. When this is done one finds that

$$I_{w1}(\infty)=I_{R1}(\infty)=(1-\delta_E)I_b\cdot\frac{(1-\epsilon^{-\gamma})}{\gamma(1+\epsilon^{-2\gamma})} \tag{21}$$

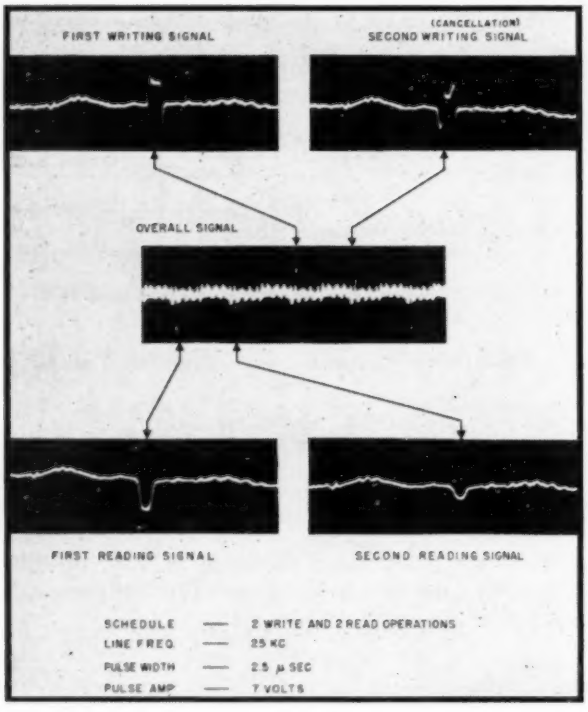
and

$$I_{w2}(\infty)=I_{R2}(\infty)=(1-\delta_E)I_b\cdot\frac{\epsilon^{-\gamma}(1-\epsilon^{-\gamma})}{\gamma(1+\epsilon^{-2\gamma})}. \tag{22}$$

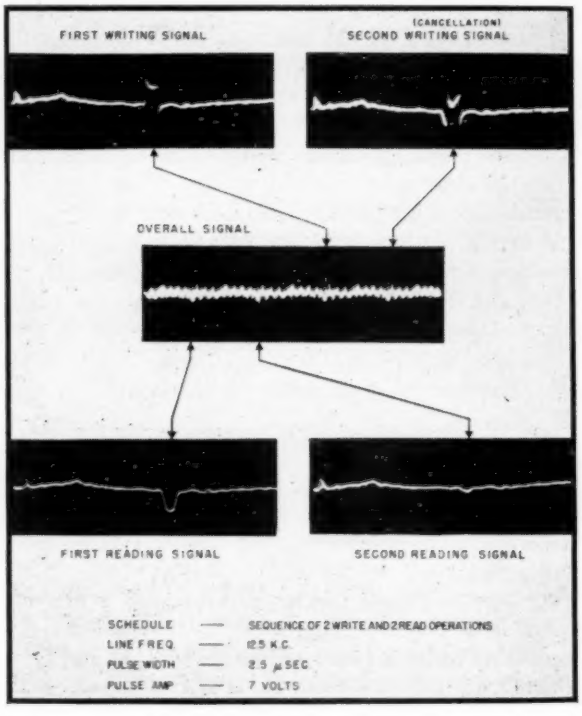
For a value of γ in the neighborhood of unity, the magnitude of A will be small, and hence the initial conditions for the first writing operation will be essentially the same as those assumed in Case I. Similarly, the basis for comparison of the second writing signal and the reading signal with Cases II and III, respectively, is also established.

Oscillograms of the writing and reading signals for a schedule as outlined above are shown (Figs. 9(a) and (b)) for $\gamma_{avg}=0.8$ and $\gamma_{avg}=1.6$. In general, a very good agreement is noted between these signals and their predicted counterparts in Figs. 5–7. It was necessary here to define the term γ_{avg} as the average of the writing and reading decrements, since for a 7-volt signal the effective secondary emission characteristic is not too linear or even symmetrical. This may be seen from Fig. 10, in which the initial amplitude of a writing signal (followed by three reading operations to insure essentially complete erasure for $\gamma\cong 1$) is taken as a measure of $(1-\delta)I_b$ and is plotted, in arbitrary units, *versus* the input voltage. The assumption of linearity is evidently valid over a range of about ± 5 volts. This is further substantiated by a plot of the *first* reading signal against the input pulse amplitude, which gives a sort of input-output characteristic for the tube. Here again it is seen that for small input signals ($<\pm 5$ volts) one may consider the operation to be approximately linear.

While these oscillograms give a fair measure of confirmation to our analysis, it is desirable to obtain more quantitative support for this theory. This may be obtained by comparing observed steady-state signal amplitudes with the theoretical amplitudes given by (21) and (22). (See Fig. 11.) It will be noted that the experimentally observed amplitudes agree fairly well with the calculated amplitudes. It should be mentioned that the sets of points were obtained by some rather inaccurate measurements of write and read signal amplitudes on a synchroscope as the average γ was changed by varying the velocity of the storage-tube sweep or scan. Furthermore, in order to provide a basis of comparison the first reading in each set was scaled to agree with the theoretical value for that particular γ and the same scale factor applied to the remaining


FIRST WRITING SIGNAL
(CANCELLATION) SECOND WRITING SIGNAL
OVERALL SIGNAL
FIRST READING SIGNAL
SECOND READING SIGNAL
SCHEDULE — 2 WRITE AND 2 READ OPERATIONS
LINE FREQ. — 25 KC
PULSE WIDTH — 2.5 μ SEC
PULSE AMP — 7 VOLTS

(a) $\gamma_{avg}=0.8$


FIRST WRITING SIGNAL
(CANCELLATION) SECOND WRITING SIGNAL
OVERALL SIGNAL
FIRST READING SIGNAL
SECOND READING SIGNAL
SCHEDULE — SEQUENCE OF 2 WRITE AND 2 READ OPERATIONS
LINE FREQ. — 125 K.C.
PULSE WIDTH — 2.5 μ SEC
PULSE AMP — 7 VOLTS

(b) $\gamma_{avg}=1.6$

FIG. 9. Storage-tube output signals.

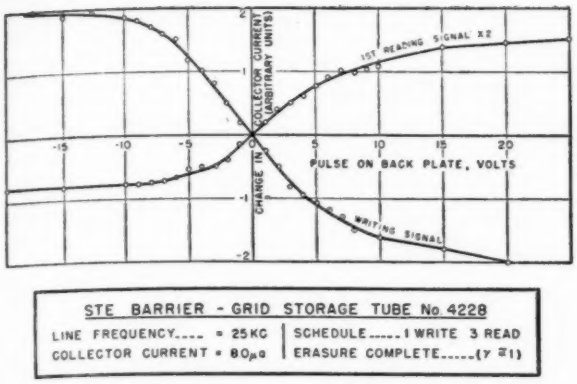

CHANGE IN COLLECTOR CURRENT (ARBITRARY UNITS)
1ST READING SIGNAL X2
PULSE ON BACK PLATE, VOLTS
WRITING SIGNAL
STE BARRIER - GRID STORAGE TUBE No 4228
LINE FREQUENCY.... = 25KC
COLLECTOR CURRENT = 80μa
SCHEDULE.....1 WRITE 3 READ
ERASURE COMPLETE.....(γ ≅1)

FIG. 10. Amplitude of writing signal and first reading signal *vs.* input pulse amplitude.

observations in the set. Thus, no comparison of absolute values is made here, but rather a comparison of the patterns or curves along which these points fall; from these results one then draws the conclusion that a rather good correlation between predicted and observed curves exists. It might be mentioned again that these amplitudes are difficult to measure with any great degree of accuracy, especially when the applied signal is kept low (at about three-volt peak) to insure good linearity since the base-line disturbance or background "noise" generated by the tube thus appears proportionately larger. Hence, the data of Fig. 11 indicate that the analysis is correct to at least a good first approximation.

Further support for the analysis is obtained by plotting experimentally observed values[9] of γ *versus* the reciprocal of the scanning velocity for various values of collector current (Fig. 12). It will be noted that the linear relationship between γ and $1/v$ is confirmed and also to a more approximate degree the direct variation of γ with I_b.

SUMMARY AND CONCLUSIONS

It will be noted that the shape of the stored voltage distribution in each of the three cases considered is the same, differing only in magnitude, the effectiveness of the erasing action in Case III being measured by the factor $\epsilon^{-\gamma}$ and the scale factor in Case II being the sum of those in I and III. This is a very useful result, since it means that for sufficiently high values of γ (say, above 2) the final value of the stored voltage distribution is dependent only upon the applied signal and is practically independent of the initial charge distribution on the dielectric. Therefore, under these conditions the signal being stored takes precedence over any previously in storage and no separate erasing operation is required.

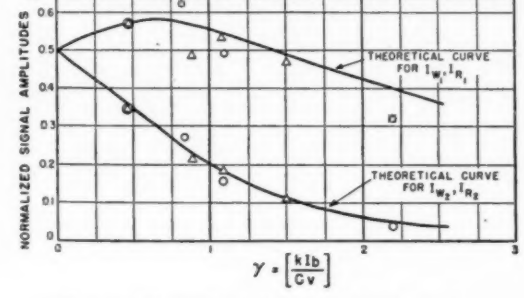

SCHEDULE: 2 WRITE - 2 READ, IN SEQUENCE
OBSERVED AVERAGE AMPLITUDES, SET A.....O
OBSERVED AVERAGE AMPLITUDES, SET B.....Δ
NORMALIZED SIGNAL AMPLITUDES
THEORETICAL CURVE FOR
THEORETICAL CURVE FOR

FIG. 11. Normalized signal amplitudes *vs.* γ.

[9] This data was obtained by T. F. Rogers and J. Dumanian.

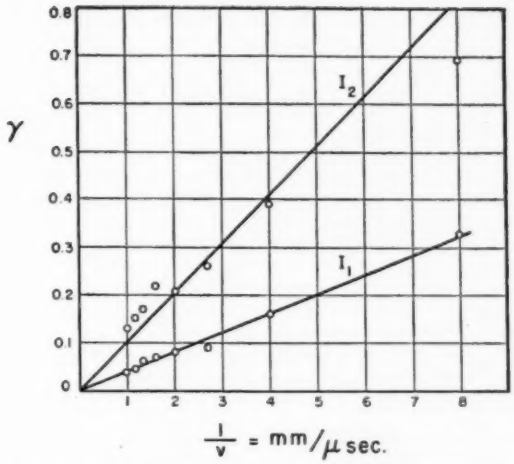


FIG. 12. γ *vs.* $(1/v)$ (for two values of collector current).

The factor $\gamma = kI_b/vC_x$ is evidently an extremely important quantity in governing the tube operation; the reason for this may be surmised from an understanding of its physical meaning. The expression for γ may be written as

$$\gamma = \frac{kI_b}{vC_x} = \frac{(1/v)}{(C_x/kI_b)} = \frac{(1/v)}{R_eC_x},$$

where $R_e = 1/kI_b$. This interpretation is dimensionally valid since k, the slope of the secondary emission curve, has the dimensions of $(\text{volts})^{-1}$; therefore $1/kI_b$ will have the dimensions of ohms multiplied by length. R_eC_x may be interpreted to be the time constant for charging the dielectric in the absence of beam motion and hence, γ is seen to be the ratio of the time it takes the beam to traverse a given spot to the time it would take the beam to charge the same spot if the beam were stationary. For reasonably successful charging action it would be expected that these times should be somewhere near equal, or that the important range of γ-values should be in the vicinity of unity; this expectation is borne out by the foregoing analysis.

ACKNOWLEDGMENTS

The author wishes to express his appreciation for the help and criticisms received in this study from many of the members of the Air Force Cambridge Research Laboratories and from Dr. A. S. Jensen of the RCA Laboratories, Princeton, New Jersey.

Graphical Representation of Particle Trajectories in a Moving Reference System

M. GARBUNY
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received April 17, 1950)

A graphical method is derived for the analysis of microwave electron tubes, ion accelerators etc., which refers particle positions and velocities to a moving reference system. If the forces are dependent on time only, the trajectories are transformed into straight lines. For inhomogeneous fields an approximation procedure applies. To demonstrate the capabilities of this method a brief treatment of the transit time phenomena in cavity triodes is outlined.

I. INTRODUCTION

THE analysis of transit time phenomena of particles under the influence of varying forces, as in electron tubes, leads often to complicated transcendental equations. Usually cumbersome graphical methods are employed to arrive at such kinematic values as arrival velocities, transit angles, and current distribution.[1] A typical example is the simple case of linear electron motion between two electrodes which produce superimposed homogeneous static and radiofrequency fields, i.e. a type of acceleration $a(t)$ described by

$$a(t)=A\sin t+B, \tag{1}$$

t being measured in electrical radians. An electron leaving the first electrode at time t_1 with a velocity v_0 arrives at a distance s at time t with a velocity v. One obtains by successive integrations:

$$v=A(\cos t_1-\cos t)+B(t-t_1)+v_0, \tag{2}$$

$$s=A[(t-t_1)(\cos t_1+v_0/A)+\sin t_1-\sin t]+B(t-t_1)^2/2. \tag{3}$$

The usual problem is then to evaluate transit times $(t-t_1)$ and velocities for a given electrode distance s with various starting times t_1 as parameter, and it is evident that in such representation elaborate graphical methods are called for such as a point-by-point construction of all the trajectories $s(t)$ with various values t_1.

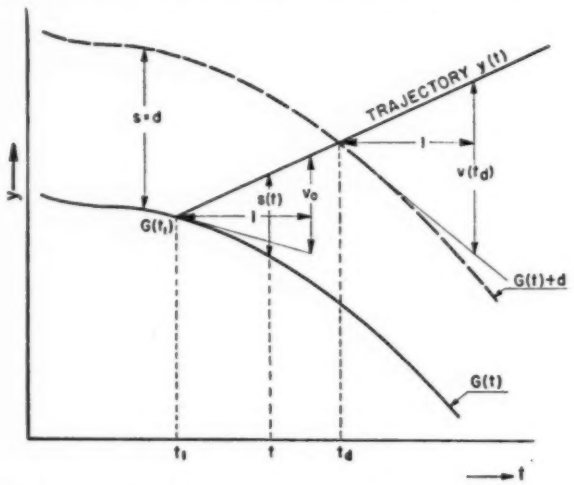


FIG. 1. Method of obtaining straight lines as trajectories.

In the following a graphical method will be described which in essence amounts to referring all trajectories to a coordinate system which moves along with one of the particles which is chosen arbitrarily for greatest simplicity of the procedure. If the force is a function only of time, as in case of homogeneous fields, all trajectories become straight lines. This means in terms of expediency that instead of constructing a great number of complicated trajectories or performing equivalent graphical operations one has to use only one relatively simple trajectory as a reference system, and all the others appear as straight lines, permitting a simple evaluation of such values as were mentioned above.

II. DERIVATION OF METHOD

Let us assume a particle enters a field at time t_1 with a velocity v_0, and the forces acting on it depend only on time so that the resulting acceleration is

$$a=f(t). \tag{4}$$

Let the motion be one-dimensional for reasons of simplicity, although the three-dimensional case can be obtained by considering the components $f_x(t)$, $f_y(t)$, $f_z(t)$ separately. We now define:

$$F(t)=-\int_p^t f(t)dt;\quad G(t)=\int_q^t F(t)dt, \tag{5}$$

where the lower limits p and q of the integrals can be chosen arbitrarily or preferably so that the simplest expressions for F and G result. One obtains then for velocity and distance reached at time t:

$$v=F(t_1)-F(t)+v_0=v_0+G'(t_1)-G'(t), \tag{6}$$

$$s=G(t_1)-G(t)+(v_0+G'(t_1))\cdot(t-t_1). \tag{7}$$

If one writes (7) in the following form as two equations

$$G(t)+s(t)\equiv y(t)=(v_0+G'(t_1))(t-t_1)+G(t_1) \tag{8}$$

one obtains two statements. The left side of (8) defines $y(t)$ as a transformed trajectory of position *vs.* time by the stipulation that

$$s(t)=y-G, \tag{9}$$

[1] A notable exception is R. Kompfner's method, Wireless Engr. **19**, 2, (1942), which for a special case overlaps with the method described here.

i.e. that the distance traveled equals the vertical distance between the trajectory and the reference curve $G(t)$. The right side of (8) states that the trajectory $y(t)$ is a straight line of slope $m=v_0+G'(t_1)$ through the point with the coordinates t_1, $G(t_1)$. The velocity at each point

$$v(t)=ds/dt=y'(t)-G'(t) \tag{10}$$

equals the difference between the slopes of the trajectory $y(t)$ and the reference curve $G(t)$. The physical interpretation of this procedure follows now easily. $G(t)$ is the trajectory of a selected particle in rectangular coordinates. With (9) and (10), one in effect refers positions and velocities to that of the selected moving particle, thereby attaining straight lines for the newly defined trajectories $y(t)$.

One thus arrives at a simple graphical method for the analysis of transit time phenomena, see Fig. 1. First the reference curve $G(t)$ is plotted, as defined by (4) and (5). Next, the straight line trajectories are drawn starting from $G(t)$ at various points t_1, $G(t_1)$. The slope of the trajectory equals the sum of the slope of the curve $G(t)$ at point t_1 and the initial velocity v_0 at this point; if the particle starts at rest, a tangent results. The intersection of the trajectory with a curve parallel to $G(t)$ at a distance $s=d$ marks the arrival time t_d. The arrival velocity is given by the vertical distance of the trajectory from the tangent to the curve at the intersection, at distance one (unit of the abscissa) from this intersection. Other kinetic values can be obtained by simple additional constructions.

A review of the literature on pertinent graphical methods revealed that for the special case $f(t)=A\sin t$ the described transformation is identical with Kompfner's procedure,[1] which however arrives at parabolic trajectories for the more general case of (1), $f(t)=A\sin t+B$. Another special case is $f(t)=0$, $G(t)=0$ for field-free drift spaces which leads to Applegate's diagram.[2]

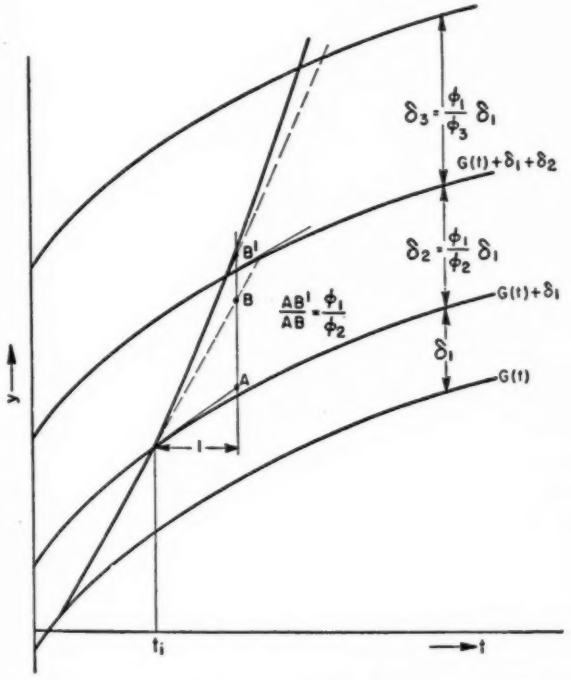


FIG. 2. Construction of trajectories in inhomogeneous fields.

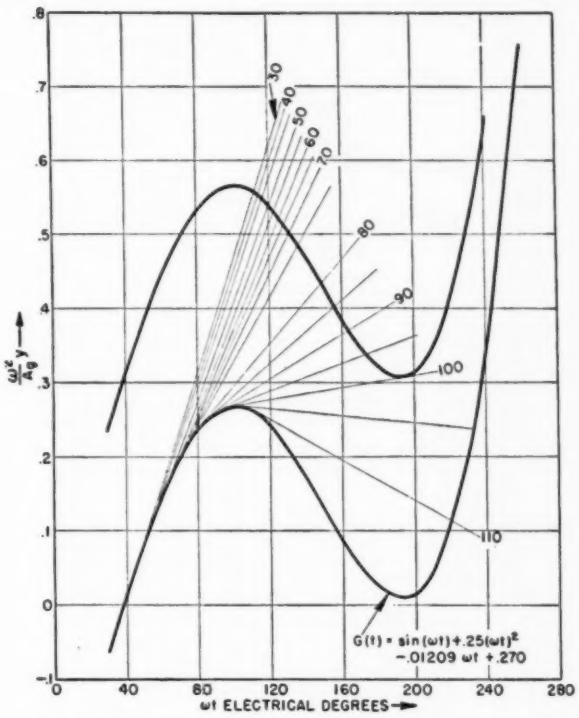


FIG. 3. Transformed electron trajectories for idealized conditions of cathode-grid cavities in dimensionless units. The chart results from the choice of $B_0/A_0=0.50$, $\delta_0=0.298$ (e.g., $d/\lambda=3.85\times10^{-3}$, $V_0=1$ Kv) and a p and q in (5) so as to increase the accuracy of the graph. The value for δ_0 corresponds to the limitation that the first emitted electron and the last not reflected electron leave at 30 and 100 electrical degrees, respectively.

III. EXTENSION OF THE METHOD TO INHOMOGENEOUS FIELDS

The method described above postulated forces which depended on time only. In inhomogeneous fields however the particle experiences forces which, in addition to being time variant, are also functions of position, i.e. one obtains for the acceleration instead of (4) the more general differential equation

$$d^2x/dt^2=\varphi(x)f(t). \tag{11}$$

Here $\varphi(x)$ is, for one-dimensional motion, the static field distribution along the path of the particle. In simple geometries $\varphi(x)$ can be computed, while in other cases it has to be determined empirically, e.g. with the electrolytic tank method. One adopts now the following approximation method for the transformation of the particle trajectories, see Fig. 2.

The traversed distance $x=d$ is subdivided into a number n of intervals Δx_λ which are equal and small enough so that within them the variation of $\varphi(x)$ from

[2] See, e.g., *Klystron*, Sperry Gyroscope Company, 1944.

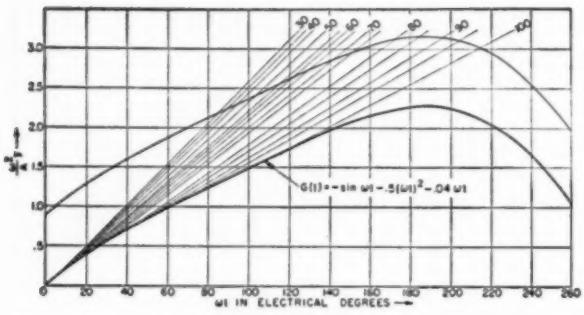


FIG. 4. Trajectories in the anode-grid cavity of a triode. Phase values marking the individual electron traces refer to the times of starting at the cathode in Fig. 3. Phase notation of the abscissa refers to the r-f voltage in the anode gap which adjusts itself to the bunched electron stream. $B/A=1$, $\delta=0.831$ (e.g., $d/\lambda=0.02$, $V=10$ Kv). A conversion efficiency of 62 percent is found by an evaluation of arrival velocities.

an average value φ_λ is sufficiently small. The condition that the forces are only time dependent is therefore approximately fulfilled within each segment.[3] The reference curves $G(t)$ and $G(t)+\delta_1$, are then plotted for the first interval Δx_1, $G(t)+\delta_1+\delta_2$ as the upper limit of the second interval Δx_2 etc. so that one plots altogether $n+1$ parallel traces of the same curve $G(t)$. The distances between the curves are inversely proportional to the fields in the intervals which they confine, i.e. $\delta_{\lambda+1}/\delta_\lambda=\varphi_\lambda/\varphi_{\lambda+1}$. The trajectories are straight lines within each interval, but at the crossing of a reference line the slope is changed. It can easily be verified that the slopes of two adjoining trajectory segments, when referred to the slope of the reference line at the intersection, are again inversely proportional to the field regions they traverse, i.e., referring to Fig. 2, $(y_{\lambda+1}'-G'(t_i))/(y_\lambda'-G'(t_i))=AB'/AB=\varphi_\lambda/\varphi_{\lambda+1}$, where $G'(t_i)$ is the slope of the reference curve at the intersection point. One obtains in this manner an approximation to the transformed trajectories in inhomogeneous fields.

[3] If the maximum variation of $\varphi(x)$ from φ_λ within Δx_λ is $d\varphi_\lambda$, the difference in the results obtained by using φ_λ and $\varphi_\lambda+d\varphi_\lambda$ is certainly larger than the actually occurring error. The difference of arrival times t_λ at the end of the interval is derived from (6) and (7) as $dt_\lambda=-d\varphi_\lambda(\Delta x_\lambda-v_{\lambda-1}(t_\lambda-t_{\lambda-1}))/v_\lambda\varphi_\lambda$. Hence the relative difference in transit time $dt_\lambda/(t_\lambda-t_{\lambda-1})\cong-bd\varphi_\lambda/\varphi_\lambda$, where $b=[\Delta x_\lambda-v_{\lambda-1}(t_\lambda-t_{\lambda-1})]/v_\lambda(t_\lambda-t_{\lambda-1})$ can be determined from the diagram. The relative difference in arrival velocity is then $dv_\lambda/v_\lambda-v_{\lambda-1})\cong d\varphi_\lambda(1-b)/\varphi_\lambda$. The actual errors will lie between zero and these upper limits.

IV. EXAMPLE: ELECTRON TRAJECTORIES IN CAVITY TYPE TRIODES

As a typical problem to which this method can be applied, the analysis of transit time phenomena in microwave triodes is briefly outlined in the following. The acceleration under negative bias conditions in the cathode-grid space is given by $f(t)=A_g\sin(\omega t)-B_g$. Using (5) and dividing by A_g/ω^2 to arrive at a dimensionless scale, one plots $G(t)$ (see Fig. 3) and a second parallel curve at a distance $\delta_g=\omega^2 d/A_g=2(2\pi d/\lambda\beta_g)^2$ wherein d/λ is the grid-cathode distance in fractional wave-lengths and β_g the fractional lightspeed of an electron corresponding to the peak r-f voltage $V_g=(m/e)A_g d$ across the gap. Since one usually chooses $B_g/A_g=0.5$ one such chart with varying parameter δ_g can describe most grid conditions which occur in the practical application of microwave tubes.

The velocities and times at which the electrons arrive at the grid are then applied as the initial conditions for the trajectories in the grid-anode space (Fig. 4). The square of the arrival velocities at the anode is proportional to the power loss, and one can in this manner, by averaging over all electrons, arrive at the total conversion efficiency.

The author wishes to express his appreciation for the many stimulating discussions with Dr. J. W. Coltman and Dr. W. Altar.

Some Observations on the Epitaxy of Sodium Chloride on Silver*

G. W. JOHNSON
Brookhaven National Laboratory, Upton, Long Island, New York
(Received May 22, 1950)

An experimental study was made of the orienting effects of silver on sodium chloride crystallized from solution. Pronounced epitaxy was observed in the following cases: (1) On oriented films of silver with (100) and also with (111) planes parallel to the surface; and (2) on polycrystalline high purity silver sheet.

INTRODUCTION

THE phenomenon of epitaxy, or oriented overgrowth of one crystal on another, has been known for over a century. The interpretation of the effect in general had to await the development of x-ray and electron diffraction techniques. Perhaps the most complete systematic study, after the x-ray lattice constants of the crystals were known, was that of Royer.[1] As a result of his work, he concluded that at least two of the following three conditions must be met before orientcd growth** can occur:

(1) If crystal A is oriented by crystal B, then A must have a two-dimensional lattice which is commensurate within limits with the exposed two-dimensional lattice of B. The lattice spacings of atoms of A and B in the mutual plane of contact do not need to be the same; the lattice spacing of one crystal may be a small whole-number multiple of that of the other.

(2) The ions of A must take positions which corresponding ions of B of the same sign would have occupied had the lattice of B been extended.

(3) A and B must have the same type of binding, e.g. both ionic crystals or both metals.

Subsequently, notable exceptions were found to these conclusions. Beukers[2] discovered that galena, a semiconductor, oriented sodium chloride, an ionic crystal. Sloat and Menzies[3] looked for oriented growth of ionic salts on silver but obtained epitaxy only with NH_4Br and CsCl. It is now well established that some metals condensed in a vacuum on cleaved ionic crystals under proper conditions will assume preferred orientations.[4–6] Among these is the growth of silver parallel to the structure of rock salt. Since the cleavage surface of rock salt is the (100) plane, the condensation of silver yields a film with the (100) orientation. Films prepared in this way have a twinned structure but after heating to 500°C for a few minutes, they are converted into relatively perfect single crystals having the same orientation as the substrate.[7]

The study of the epitaxy of ionic crystals deposited from solution on metals has not been given much attention. Brück,[5] and Goche and Wilman,[7] after removing the silver film from rock salt by solution of the latter in water, observed interference points in the diffraction patterns which were attributed to the presence of sodium chloride on the surface, which had not been removed by their rinsing procedures. The indicated orientation was that the (100) plane of sodium chloride was parallel to that of silver, but that the [100] direction of the salt was along the [110] direction of silver. The films apparently were not examined under a microscope. In a recent general discussion of crystal growth by G. P. Thomson[8] and an excellent survey of the literature by van der Merwe,[9] no examples were given of epitaxy between ionic crystals and metals other than those referred to here.

This paper is a description of some observations of marked epitaxy of sodium chloride on oriented films and on polycrystalline sheets of silver. It was found that orientation resulted even when the difference in atomic spacings for the two planes in contact was large. It was concluded that the orienting influence was exerted in the embryonic stages of growth.

EXPERIMENTAL PROCEDURE

A few drops of sodium chloride solution were placed on the surface of a suitable silver specimen (the preparation of which is described below), and after evaporation of the solvent, the disposition of the sodium chloride crystals was examined with a microscope.

Two types of silver surfaces were employed: oriented thin films and polycrystalline sheet. The use of thin films provided a ready means of obtaining surfaces with known crystallographic planes exposed. It was essential to have such surfaces in order to relate the observed orientation of the deposited sodium chloride crystals with that of the silver. Polycrystalline sheet was used to find in a qualitative way how universal the orienting influences were, and to determine whether the grain boundaries exerted any specific effects.

The method of Brück and others[4–6] was applied to the

* Research carried out at Brookhaven National Laboratory, under the auspices of the AEC.
[1] L. Royer, Bull. Soc. Franc. Miner. **51**, 7 (1928).
** Epitaxy and oriented overgrowth will be used synonymously.
[2] M. C. F. Beukers, Rec. Trav. Chem. **58**, 435 (1939).
[3] C. A. Sloat and A. W. C. Menzies, J. Phys. Chem. **35**, 2005 (1931).
[4] H. Lassen and L. Brück, Ann. Phys. (Leipzig) **22**, 65 (1935).
[5] L. Brück, Ann. Phys. (Leipzig) **26**, (5), 233 (1936).
[6] O. Rudiger, Ann. Phys. (Leipzig) **30** (5), 505 (1937).
[7] O. Goche and H. Wilman, Proc. Phys. Soc. **51**, 625 (1939).
[8] G. P. Thomson, Proc. Phys. Soc. **61**, 403 (1948).
[9] J. H. van der Merwe, Disc. Faraday Soc., "Crystal Growth," No. 5, 201 (1949).

production of films with the desired orientations. Silver was vaporized in a vacuum from a tungsten filament and condensed (1) on cleaved rocksalt, and (2) on mica. To assure that properly oriented films are formed with this technique, it is necessary that the crystals be cleaved just prior to their introduction into the vacuum chamber. In addition, both the mica and the rocksalt must be held at a temperature above 150°C while the silver is being condensed. In our work the crystals were heated to 190°C during deposition. In this way, a silver film with the (100) plane exposed was produced on rock salt, and on mica the (111). Sufficient silver was evaporated to get a film thickness of ~1000A. The films were allowed to cool in a vacuum before removing them from the tube.

To reduce the possibility that the substrate might in some way influence the observations, the rock salt was dissolved in water. The film floated on the surface of the water and was readily picked up on a polished silver plate. The epitaxy experiments were made on the surface of the film that had not been in contact with the salt.

The film on mica was ready for testing for epitaxy as soon as it was taken from the vacuum chamber. It was unnecessary to remove the film from the substrate since it was shown by Royer[10] that mica does not orient sodium chloride. The preparation of the polycrystalline silver sheet involved cleaning, heat-treatment, and in some cases etching. Square pieces ($\frac{1}{2}'' \times \frac{1}{2}''$) were cut from a 20-mil thick plate. These squares were washed carefully in detergent and rinsed successively in distilled water, benzene, and alcohol. The specimens were heated in a vacuum at 700°C for one hour and furnace cooled to removed the effects of the rolling of the sheet during its fabrication. Relatively large grains (~0.5 mm in diameter) resulted from this treatment. Each specimen then was electrolytically polished using published methods.[11,12] Although polishing was not essential to the production of epitaxy, a polished surface was easier to examine. To reveal the grain boundaries some specimens were etched in a mixture of equal parts of five percent solutions of potassium cyanide and ammonium persulfate.

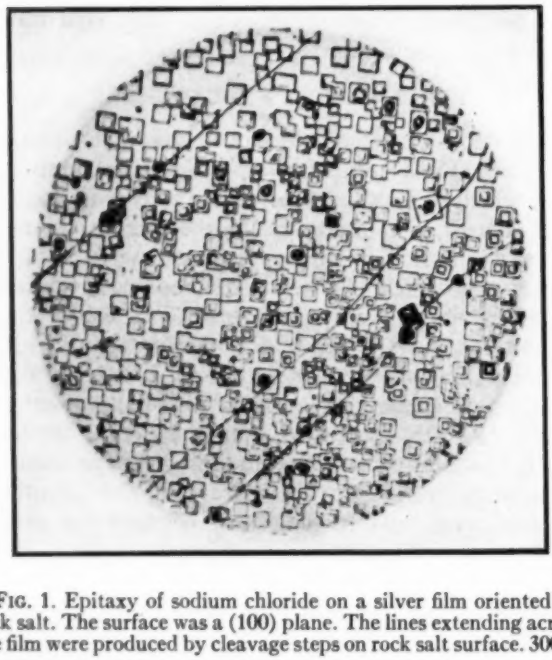


FIG. 1. Epitaxy of sodium chloride on a silver film oriented by rock salt. The surface was a (100) plane. The lines extending across the film were produced by cleavage steps on rock salt surface. 300×

[10] L. Royer, Comptes Rendus **194**, 620 (1932).

The sodium chloride was crystallized from saturated solutions in mixtures of water and ethyl alcohol. By varying the amount of alcohol the rate of evaporation could be changed permitting the control of crystal size over wide limits. If large crystals were desired, an aqueous solution was necessary. On the other hand, if extremely small crystals were required, a strongly alcoholic solution was used.

To test for epitaxy, a few drops of the solution were placed on the surface. The surface was then dried under an electric hair-dryer. On a grease-free surface, the crystallization was spread uniformly over the surface except for variations from grain to grain due to orientation effects.

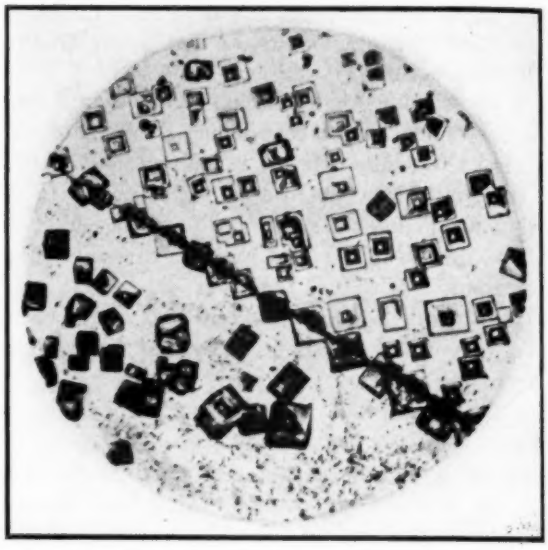


FIG. 2. Sodium chloride crystals near edge of silver film oriented by rock salt. Silver film supported by a polycrystalline silver plate. Note growth of crystals from edge of film. 300×

Each specimen was then examined under a microscope to observe the presence or absence of preferred orientations. If the crystal size was unfavorable, the surface was cleaned by rinsing in distilled water. Another solution then was tried to get the desired crystal size. This procedure could be repeated many times, although usually after several rinses the surface would show evidence of contamination, as indicated by failure of the solution to wet the surface completely. Representative

[11] Shuttleworth, King, and Chalmers, Metal Treatment **14** (No. 51) 161 (1947).

[12] L. I. Gilbertson and O. W. Fortner, Trans. Electrochem. Soc. **81**, 199 (1942).

arrangements were photographed using a Bausch and Lomb metallograph under bright field vertical illumination.

Material

The silver was of high purity (99.999 percent) supplied by Sigmund Cohen. The rock-salt crystals were cleaved from a synthetic single crystal from the Harshaw Chemical Company. The mica was ordinary muscovite.

RESULTS

Thin Films—On the films with the (100) plane exposed, almost complete orientation of the sodium chloride was achieved. Characteristic results are shown

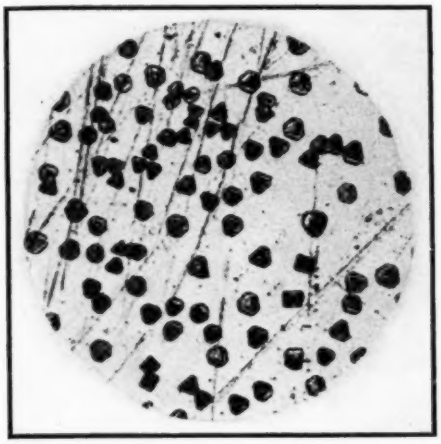


FIG. 3. Sodium chloride crystallized on silver film oriented by mica. 400×

in Fig. 1. The crystals formed rectangular pyramids whose bases were in contact with the silver surface. The straight edges forming the sides of the squares were [100] directions of sodium chloride and were directed along the [110] of the underlying silver. Thus the (100) planes of the salt and the silver were parallel, but the cube edge of the salt was aligned with the face diagonal of the silver.

Since the silver film was mounted on a polycrystalline silver plate, it was of interest to compare the areas of the film and the plate adjacent to the edge of the film to determine if the plate was affecting the results. Such an edge is shown in Fig. 2. On top of the film proper the sodium chloride crystals were oriented as illustrated in Fig. 1, whereas on the plate no general preferred alignment was noted. The edge of the film was the [100] direction of silver. The rotation of the salt crystals by 45° with respect to cube edge of the silver is plainly indicated. Of additional interest is the rather extended growth of crystals, which had nucleated at the edge of the film, over the support plate. In these cases, the crystal form was maintained even though the growth took place over a polycrystalline substrate. This means that the crystal form was determined in the embryonic

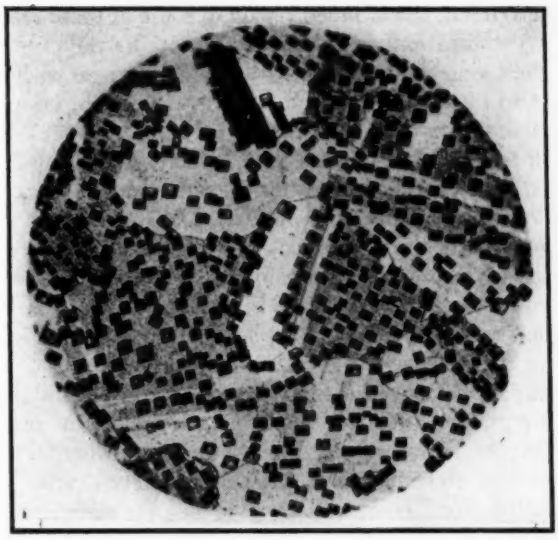


FIG. 4. Sodium chloride crystals on polycrystalline silver. Electropolished and etched. Note variation of density of crystallization from grain to grain and between twinned portions of grains. 150×

stage of development and subsequent growth was unaffected by the plate. This point will be amplified in the discussion.

On the silver film deposited on mica, the alignment was strict. The sodium chloride crystals assumed triangular pyramidal shapes. The results are presented in Fig. 3.

Striking effects were noted with polycrystalline silver. Practically complete epitaxy occurred for silver prepared in any of the following manners: electropolished and etched; electropolished but not etched; neither polished nor etched. In all cases, the surface was cleaned and heat-treated at 700°C for one hour prior to the application of the sodium chloride solution. The nature of the oriented overgrowth was the same for each of the three methods of surface preparation.

Characteristic results for a polished and etched specimen are shown in Fig. 4. The grain and twin boundaries may be seen in the field. The twinned regions are the generally banded areas within a grain. Each part of the grain which is twinned with the same orientation etches in the same way. Twinning is rather general over the whole area. Silver is well known to have a propensity for this behavior.

Those areas within a grain which etch in the same way are known to have the same orientation. Inspection of Fig. 4 reveals that the sodium chloride crystals had the same orientation over those parts of the grain which had the same appearance after being etched. In the dark areas of the large grain in the center of the figure, all the crystals had the same orientation. The white areas, which were devoid of crystals, were twinned regions of that grain.

The crystals on neighboring grains were oriented in various directions but on each grain one orientation was

preferred. On the twinned region of some of these grains no crystallization was observed. On others the crystals on the twins were all rotated through an angle with respect to those on the rest of the grain. Several examples of these effects may be seen in Fig. 4.

On polished but unetched silver similar behavior was noted. The crystals of sodium chloride clearly indicated the location and size of the grains. The similarity of the appearance of the sodium chloride crystals shown in the center of Fig. 5 and those in Fig. 1 suggest that the orientation of the substrate in the two cases was the same. Presumably the (100) plane of the grain in Fig. 5 lay in the surface of the plate.

It might have been anticipated that the discontinuities at the grain boundaries would provide places of ready nucleation. However, the grain boundaries, whether the silver was etched or not, were not preferred sites for crystallization. There was no evidence for such preference on any of the surfaces examined.

The density of crystallization varied from grain to grain. On some grains, in fact, no crystallization took place. The variation in density of crystallization was applied in one case to reveal the grain structure. A strongly alcoholic solution dried rapidly on a polished

FIG. 5. Sodium chloride crystals on polycrystalline silver sheet. Electropolished and unetched. 150×

surface in a current of warm air was used to produce crystals ~1 micron in diameter. The variations in the density of crystallization produced sufficient contrast to outline the grains (Fig. 6). The same area was then rinsed and etched to delineate the grains in the usual manner. The result of this is shown in Fig. 7. The faithfulness of the reproduction by crystallization is apparent through a comparison of Figs. 6 and 7.

Another observation of interest was the growth of crystals across grain or twin boundaries without modification. From this it appeared that the form of the crystal was determined in the early stages of growth. Many examples of crystals which had nucleated on one side of a boundary and grown across may be found in Figs. 4 and 5. Additional illustrations were furnished when caesium chloride instead of sodium chloride was crystallized on the surface. The crystals grew farther into adjoining grains (Fig. 8) than did sodium chloride crystals. These results suggest that the influence of the substrate was exerted only in the embryonic stages of crystallization. Once the crystal had nucleated, subsequent growth was not affected by the nature of the

FIG. 6. Grain structure of polished polycrystalline silver revealed by variation of density of crystallization on the surface. 150×

underlying material. Further support to this view was given by the growth of crystals from the edge of the thin film shown in Fig. 2.

A silver surface, once it had been prepared so that it was capable of orienting sodium chloride, retained this capacity for months. There was no evidence of any aging effects.

DISCUSSION

There are four general features of the results which are of interest: (1) Pronounced epitaxy was found to exist between a metallic surface and an ionic crystal; (2) each area of the substrate of a given orientation was characterized by a particular orientation of the overgrowth; (3) the density of crystallization was dependent upon the orientation of the grain; (4) the crystals of sodium chloride often grew across grain and twin boundaries without modification. Each of these points is discussed in the following.

At the outset, it is clear that the third condition suggested by Royer,[1] requiring that the overgrowth and the substrate have the same type of binding, must be either modified or abandoned. Here we have strong evidence of epitaxy between a substrate with metallic binding and

an ionic crystal. The only question which might arise in interpretation is whether the salt was oriented directly by the substrate or by a compound on the surface. In particular the presence of an oxide layer would be suspected. Since each grain of silver in the polycrystalline specimens produced either a unique alignment or none, it follows that if an oxide layer were present it bore a definite crystallographic relationship to each grain of the substrate. While there are many such examples of oriented growth of oxides on metals, none have been reported for silver. Actually silver oxide is quite unstable

FIG. 7. Same area of polycrystalline silver sheet shown in Fig. 6 after etching. 150×

above 200°C and its rate of formation is exceedingly low at lower temperatures. As a result of an extensive investigation of the stability of silver oxide (Ag_2O) and the adsorption of oxygen on silver, Benton and Drake[13] concluded that silver acted upon by oxygen at room temperature becomes covered with a monomolecular layer. The rate of oxide formation would be negligibly small. From photoelectric work function measurements on silver, Brewer[14] found that an oxidized surface could be restored to a value characteristic of clean silver by heating at 600°C for several hours under vacuum. It will be recalled that the polycrystalline silver used in the present work was baked in a vacuum at 700°C for one hour and cooled in the furnace; and that the silver films were prepared by vacuum distillation. Consequently, while an adsorbed layer of oxygen probably existed on the surface, it is unlikely that an oxide film was present.

The results for polycrystalline metal and for oriented films indicated that the fit of the two lattices in contact need not be close for epitaxy to occur. On polycrystalline silver some form of alignment was produced on almost every grain. The universality of the effect in itself suggests that closeness of fit cannot be a strict requirement. Quantitative justification of this conclusion may be found in the results for oriented films.

When crystallized on a (100) plane of silver, the [100] direction of sodium chloride was directed along the [110] direction of the metal. The distance between adjacent atoms along the [100] axis of sodium chloride is 2.81A, and that between consecutive silver atoms along the [110] is 2.88A. Thus, with this arrangement the difference in atomic spacings 2.5 percent. As pointed out by G. P. Thomson[8] this relative orientation is the one of closest fit.

For (111) planes in contact, which was the situation for the film oriented by mica, the differences in atomic spacings of the sodium chloride and the silver lattice amounted to 73 percent. The existence of epitaxy between silver and sodium chloride, therefore, does not depend markedly on closeness of fit of the two lattices. Sloat and Menziers[3] obtained oriented growth on silver for only those salts whose lattice parameters differed from those of the metal by less than one percent; and, they concluded that the lattices must be commensurate

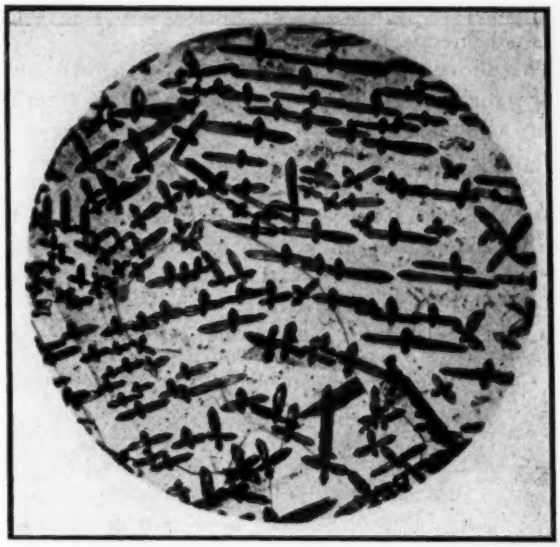


FIG. 8. Epitaxy of caesium chloride on polycrystalline silver. Note extensive growth of crystals across grain and twin boundaries without modification. 150×

within that tolerance. However, the present results show a much greater difference in lattice dimensions can be tolerated.

The observation that the density of crystallization depended upon the orientation of the substrate might have been anticipated. For electrodeposited metals, Finch and Sun[15] found that the adherence between the coat and the underlying surface was greatest when strong epitaxy existed between the two metals. Royer[1] noted that when a mixture of glass, mica, gypsum,

[13] A. F. Benton and L. C. Drake, J. Am. Chem. Soc. **56**, 255 (1934).

[14] A. K. Brewer, J. Am. Chem. Soc. **56**, 1909 (1934).

[15] G. I. Finch and C. H. Sun, Trans. Farady Soc. **32**, 852 (1936).

calcite, talc, and orpiment were placed in a solution of ammonium iodide, the crystals always appeared first on the mica. The presence of epitaxy implies strong interatomic forces between the crystallizing atoms and those of the substrate. Therefore, a more rapid nucleation would be expected to take place on those surfaces favoring oriented growth. It is of interest to note that while contact of a crystal of a solute with a solution is the only means of completely preventing supersaturation, a surface which can show epitaxy with the solute may prevent a degree of supersaturation otherwise possible.[16]

For the silver film prepared on rock salt it was possible that nuclei were mechanically retained in the film after its removal from the salt. Such nuclei would serve as centers of crystallization when the film was tested for epitaxy. The growth, in this event, would have been parallel to the original rock salt lattice. Even after thorough rinsing a few crystals had this orientation, but practically all were rotated 45° with respect to this direction. Thus, the observed effects were undoubtedly characteristic of the silver film.

On the film with the (111) plane produced on mica, the observed results must be attributed to the silver, since Royer[10] showed that sodium chloride is not oriented directly by mica.

That the orientations of the crystals of the overgrowth were determined in the embryonic stages of crystallization was indicated by two observations: (1) the growth of crystals from the edges of an oriented film, and (2) the growth of crystals across grain or twin boundaries. Thus, the crystals often extended distances of ~0.01 mm, or several thousand atomic diameters, over areas of the substrate having an orientation different from that at the point of initial formation of the crystal. The fact that in many instances, a major portion of the crystal was resting on a surface which favored another orientation produced no modification of the overgrowth. More examples of this effect were found also with caesium chloride crystals on polycrystalline silver (Fig. 8). This situation could prevail only if the orientation of the crystal were determined on an area of much smaller dimensions than those of the final crystal. From the extent of growth across boundaries, a rough upper limit may be assigned to the size of the embryo. In many cases the crystallites were nearly centered on boundaries. Since the crystal diameters were about 0.01 mm, the embryo was certainly less than one-tenth of this value. Indeed, as pointed out by G. P. Thomson[8] and van der Merwe,[9] it is difficult to understand how the embryo could be larger than a few atomic diameters.

The epitaxy of sodium chloride provided a ready means of testing the degree of orientation of thin films without resorting to electron diffraction. Electron diffraction measurements are required to establish the crystallographic relationships of the overgrowth to the substrate, but once these are known, a rinse in sodium chloride solution suffices as a check. The rinsing procedure gives a representation of the orientation of the entire film, while the usual electron diffraction arrangement permits the examination of only a relatively small area. It must be emphasized however, that observations of epitaxy alone cannot be used to determine the orientation of the substrate, but can be used as a check after the expected orientation is known.

[16] This was pointed out by Dr. W. Rubinson. Private communication.

ACKNOWLEDGMENTS

The author is indebted to Dr. William Rubinson for a review of the literature and for many helpful criticisms; to Dr. David Gurinsky for constructive discussions; to Messrs. O. Kammerer and A. Cendrowski for the photomicrography; and to Mr. N. Bylock for technical assistance. All are members of the staff of the Brookhaven National Laboratory.

Letters to the Editor

Note on Stability of Electron Flow in the Presence of Positive Ions

J. R. PIERCE
Bell Telephone Laboratories, Murray Hill, New Jersey
July 17, 1950

THE writer once calculated what he believed to be the limiting stable current in parallel electron flow normal to parallel conducting bounding planes (as, grids).[1] The conditions assumed were the following: Two parallel grids are spaced L cm apart. A current i_0 amp./cm^2 of electrons with a velocity u_0 specified by the accelerating voltage V_0 is injected through one. In the space between the grids, a positive charge density of immobile ions is assumed, such that if the electrons retain their velocity of injection the net space charge between the grids will be zero. The question is, for how large a value of i_0 will the flow be stable? The answer was sought by considering a small sinusoidal perturbation of the flow of frequency ω. It was found that if

$$\frac{\omega_0 L}{u_0}=\pi+\alpha \tag{1}$$

where α is a positive quantity small compared with π, the boundary conditions are satisfied for a frequency

$$\begin{aligned}\omega&=-j\pi u_0\alpha/4L\\ j\omega&=\pi u_0\alpha/4L.\end{aligned} \tag{2}$$

Here ω_0 is the electron plasma frequency.

$$\omega_0^2=\frac{(e/m)\rho_0}{\epsilon} \tag{3}$$

where e/m is the charge-to-mass ratio of the electron, ρ_0 is the electronic charge density and ϵ is the permittivity of space. Thus, for $\omega_0 L/u_0$ a little greater than π, perturbations in the flow grow exponentially with time; the flow is unstable.

The situation analyzed was a little unrealistic in that the ions were assumed not to move. Under these circumstances, one obtains an expression for the phase constants of the space-charge waves

$$\beta=\frac{\omega}{u_0}\pm\frac{\omega_0}{u_0}. \tag{4}$$

If we allow motion of the ions, we find[2]

$$\beta=\frac{\omega}{u_0}\pm\frac{\omega_0}{u_0}\left(1-\left(\frac{\omega_i}{\omega}\right)^2\right)^{-\frac{1}{2}}. \tag{5}$$

Here ω_i is the ion plasma frequency

$$\omega_i^2=\frac{(q/m)\rho_i}{\epsilon} \tag{6}$$

where q/m is the charge-to-mass ratio of the ions and ρ_i is the ion charge density (assumed to be equal and opposite to the electron charge density).

We can use all the relations developed in reference 1 if we there replace ω_0 by ω_1, where

$$\omega_1=\omega_0[1-(\omega_i/\omega)^2]^{-\frac{1}{2}}. \tag{7}$$

Suppose, for instance, that we let

$$\frac{\omega_1 L}{u_0}=\pi+\alpha. \tag{8}$$

We now assume that α is small compared with π, so that ω is given in terms of α by (2). Then, using (7) and (8), we obtain

$$\alpha^2\left(\frac{\omega_0 L}{u_0}\right)^2=(\pi+\alpha)^2\left[\alpha^2-\left(\frac{\omega_i L}{u_0}\right)^2\left(\frac{4}{\pi}\right)^2\right]. \tag{9}$$

We have already assumed that $\alpha\ll\pi$, so α will be neglected in comparison with π in the parentheses. This gives for α

$$\alpha=\pm\frac{4(\omega_i L/u_0)}{[(\omega_0 L/u_0)^2-\pi^2]^{\frac{1}{2}}}. \tag{10}$$

We note that for

$$\frac{\omega_0 L}{u_0}<\pi \tag{11}$$

α will be imaginary, the frequency will be real and the flow will be stable. For

$$\frac{\omega_0 L}{u_0}>\pi, \tag{12}$$

α will be real and positive for one root and hence $j\omega$ will be a real positive quantity and the beam will be unstable.

This argument does not apply near

$$\frac{\omega_0 L}{u_0}=\pi \tag{13}$$

since α will be large in this vicinity. We note, however, that for equal electron and ion charge densities

$$\frac{(\omega_i L/u_0)}{(\omega_0 L/u_0)}=\left(\frac{\text{mass of electron}}{\text{mass of ion}}\right)^{\frac{1}{2}}. \tag{14}$$

Thus, for massive ions, α will still be small compared with π for $(\omega_0 L/u_0)$ fairly near to π.

The conclusion then seems to be that even when the positive charge consists of massive ions rather than a fixed charge, there is a change from a stable to an unstable condition near (13), which represents the value of the electron plasma frequency ω_0 for which the flow becomes unstable for a fixed positive charge. It is of course possible that some instability for a lower value of ω_0 has been overlooked.

This note is to be construed merely as an amplification and clarification of work published earlier. Experiments seem generally to indicate that electron flow becomes unstable for currents nearer to those calculated assuming no positive ions present than to the larger currents calculated for the presence of ions.

[1] J. R. Pierce, J. App. Phys. **15**, 721–726 (October, 1944).
[2] J. R. Pierce, J. App. Phys. **19**, 231 (1948).

A Modification of the Radiofrequency Mass Spectrometer

A. F. HENSON
Imperial Chemical Industries Ltd., Butterwick Research Laboratories, The Frythe, Welwyn, Herts, England
April 26, 1950

THE resolution of a single stage of the radiofrequency mass spectrometer described by Bennett[1] is poor, and in order to obtain sufficient resolution a three-stage instrument is used. A possible method of obtaining the necessary resolution without the use of so many grids in the tube would be to apply a square wave form alternating potential to a single stage, the two halves of which were separated by constant potential "velocity selection" section. The tube would consist of the ion source as in Bennett's tube, four grids and a collector plate. For the analysis of positive ions, fixed potentials (negative with respect to the ion source) would be applied to the four grids. Grids 2 and 3 would be connected together and to them would also be applied the square wave form consisting of single complete cycles (positive half followed by negative) separated by intervals. The spaces between grids, the square-wave frequency and the intervals between cycles, would be arranged so that an ion of the selected mass passing through grid 1 at the start of a negative half-cycle passes through grid 2 at the end of the negative half-cycle. During the interval between cycles it travels in the field-free space between grids 2 and 3, emerges through grid 3 at the start of the next positive half-cycle, and through grid 4 at the end of this positive

half-cycle. It then passes on to the collector plate to which a blocking potential is applied in order to collect only the most energetic ions. It is thus accelerated by the square-wave field for the whole of the time it is between grids 1 and 2 and between grids 3 and 4.

A heavier ion than the selected one travels too slowly to gain maximum energy in either half-cycle. A lighter ion can gain maximum energy from either half-cycle but not from both. Thus ions of the selected mass gain more energy than any others. Different masses would be selected by varying the square-wave frequency and interval length, keeping the ratio of period of one cycle to interval constant.

A solution for the motion of the ions in the square-wave field shows that the discrimination between ions of the selected mass m (atomic weight units) and the next heavier ion $m+1$ is much larger than between m and $m-1$. The controlling consideration is then the discrimination between m and $m-1$. In the design of such an instrument there are many variables available. The values chosen will depend on the use to which the instrument is to be put and will be a compromise between conflicting requirements. Calculations have been made illustrating the use of the instrument for work in the low mass range up to about mass 50, such as might be used for stable isotope tracer work with carbon nitrogen and oxygen. The steady potential on the grids was -200 v and the amplitude of the square wave 60 v. It was considered that an energy discrimination between neighboring masses of at least 2 ev would be necessary and that, from consideration of ease of production of the square wave form, its period should not be less than 2 μsec. Under these conditions the tube length is about 80 cm and the energy discrimination between m and $m-1$ varies between 9.6 ev for $m=13$ and 2.4 ev for $m=50$. If it were convenient to produce the square wave form with greater amplitude or shorter period, or if a smaller energy discrimination were found to be sufficient, the tube could be considerably shortened.

[1] W. H. Bennett, J. App. Phys. 21, 143 (1950).

Remark on an Aberration-Phenomenon in Electrostatic Lenses. "Kidney Effect"

E. Baş-taymaz
Swiss Federal Institute of Technology, Department of Applied Physics, Zürich, Switzerland
June 7, 1950

In a letter,[1] S. Harrison described an aberration phenomenon in electrostatic lenses, the so-called "Kidney Effect" and gave an explanation. We had observed the same phenomenon in a special electron gun structure a long time ago, not being able to give any explanation. We give here a short description of our observation.

The scheme of the electron gun is shown in Fig. 1. It consists of a tungsten-Bolzen-cathode[2] and a focusing system similar to the description given by R. R. Law. The surface of emission can be imaged sharply on a fluorescent screen. However, with this

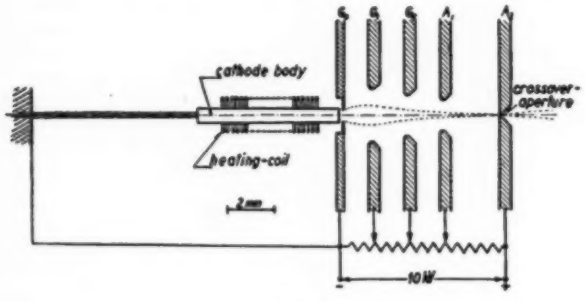


FIG. 1.

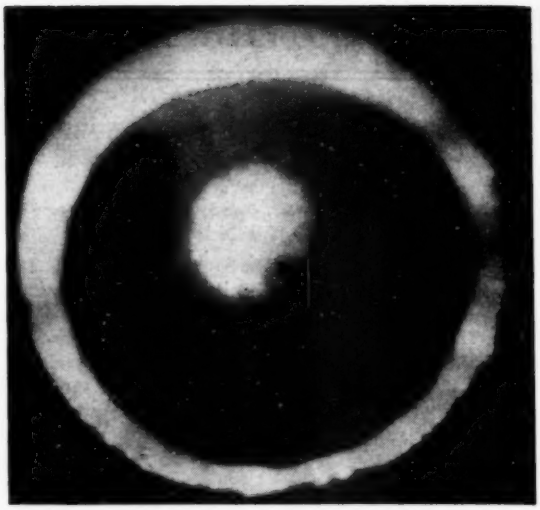


FIG. 2.

well-defined shape of the image produced by changing the strength of the lens by means of potentials of grids 1 and 2, a bright fluorescent spot occurred on the surroundings by further variation of the potentials, which grows to an illuminated circular area as shown in Fig. 2. The picture is the same as given by Harrison with only one difference that we have a greater symmetry resulting from the exact form of the electron gun. The picture, however, does not show exact symmetry and it shows how difficult it is to realize the condition of axial symmetry in electrostatic lenses.

The assumptions of Harrison make it possible to understand the phenomenon of the "Kidney Effect" in this special case, because we inserted *a priori* a little aperture in the focus (cross-over-aperture) which might cause the effect. Furthermore, we can see that this electron gun structure (Fig. 1) is showing a rather serious spherical aberration; a fact which is in accordance with our experience by which it has been ascertained that it is not possible to obtain the theoretical value of cross-over current density.

[1] S. Harrison, J. App. Phys. 20, 412 (1949).
[2] E. Baş-taymaz, Ph.D. thesis, E.T.H., Zurich, 1949 (to be published).

Dielectric Constants of Five Gases at 9400 Megacycles per Second*

James A. Murphy and Richard C. Raymond
Department of Physics, The Pennsylvania State College, State College, Pennsylvania
June 23, 1950

THE dielectric constants of five gases have been measured at pressures up to one atmosphere by a resonant cavity method similar to that reported by Jen.[1] The cavity was a copper-walled cylinder with movable plunger and temperature control, and was operated in the TE01 circular mode. Measurements of the frequency shift and Q were taken from an oscillographic display of the effective resonance curve with the help of a coupled microwave circuit. Modulation power for measurement of frequency shifts was generated by a General Radio Type 1001A signal generator and could be varied in frequency from 5 kc/sec. to 50 Mc/sec. The minimum observable change in frequency was 0.25 Mc/sec.

The dielectric constants of the gases, as observed from the resonant frequency shift at a pressure of 730 mm mercury, are listed in Table I.

TABLE I. Dielectric constants of gases computed from observed resonant frequency shift at a pressure of 730 mm of mercury and wave-length 3.19 cm.

Gas	Dielectric constant $(\epsilon'-1)\times10^3$	Temperature
Acetone	17.1	75° C
Ammonia	6.8	26°
Carbon dioxide	1.6	26°
Carbon tetrachloride	3.9	85°
Water vapor	6.6	102°

Acetone and ammonia showed appreciable broadening of the resonance curves at the maximum pressures investigated. The loss factors were computed to be 2.7×10^{-4} and 4.7×10^{-4}, respectively, at the pressures and temperatures indicated in Table I.

The results are in general agreement with those reported by others[2-6] with the exception of CO_2. The CO_2 used here was prepared by sublimation of commercial dry ice, and may have included some water vapor.

* This paper is from a thesis by James A. Murphy submitted in partial fulfillment of the requirements for the degree of Master of Science at the Graduate School of The Pennsylvania State College.

[1] C. K. Jen, J. App. Phys. **19**, 649–653 (July, 1948).
[2] C. M. Crain, Phys. Rev. **74**, 691–693 (1948).
[3] L. G. Hector and D. L. Woernley, Phys. Rev. **69**, 101 (1946).
[4] J. E. Walter and W. D. Hershberger, J. App. Phys. **17**, 814–822 (1946).
[5] C. D. Hodgman, *Handbook of Chemistry and Physics* (Chemical Rubber Publishing Company, Cleveland, Ohio, 1937), p. 1451.
[6] J. A. Saxton, Rep. Phys. Soc., London, 215–238 (1947).

A Simplified Air Flow Method for the Determination of the Surface Area of Powders

GERARD KRAUS AND JOHN R. THIEM
Applied Science Research Laboratory, University of Cincinnati, Cincinnati, Ohio

IN a recent paper appearing in this Journal, F. A. Schwertz[1] reported a series of surface area measurements on sintered Pyrex diaphragms by means of a gas flow method similar to those of Rigden[2] and Arnell.[3] A particularly useful modification of this method which is quite rapid and requires only the simplest kind of equipment has been used by the writers to measure the surface areas of powders ranging from relatively coarse granules to colloidal carbon blacks.

The specific flow rate in moles/sec. per unit pressure difference under combined Poiseuille and slip flow is, in the notation of Schwertz (reference 1, Eq. (15)):

$$\dot{G}=\frac{\epsilon^3AP_m}{k_v(1-\epsilon)^2S^2\mu LRT}+\frac{\epsilon^2\pi A}{k_d(1-\epsilon)SL(2\pi MRT)^{\frac{1}{2}}}\cdot\left(\frac{2}{f}-1\right). \quad (1)$$

Now consider a partially evacuated reservoir connected to the atmosphere through a plug of packed powder whose surface is to be measured. If the atmospheric pressure is P_0, the pressure inside the reservoir is $(P_0-\Delta P)$ and its volume V, it can easily be shown that

$$\frac{d(\Delta P)}{dt}=\frac{A\epsilon^3}{10V\mu L(1-\epsilon)^2S^2}(\Delta P)^2-\left[\frac{A\epsilon^3P_0}{5V\mu L(1-\epsilon)^2S^2}+\frac{A\epsilon^2Z}{(1-\epsilon)VSL}\left(\frac{\pi RT}{2M}\right)^{\frac{1}{2}}\right]\Delta P \quad (2)$$
$$=\beta(\Delta P)^2-\alpha\Delta P$$

where k_v has been taken as 5.0 and $(1/k_d)(2/f-1)$ has been replaced by Z. Dividing through by ΔP it follows that

$$\frac{d[\ln(\Delta P)]}{dt}=\beta\Delta P-\alpha \quad (3)$$

so that as ΔP approaches zero the slope of a plot of $\ln(\Delta P)$ *versus* t will be α.

TABLE I. Comparison of modified air flow method with established methods of surface area measurement.

	Surface area m^2/g		
	Air flow	Check	Method of check
Microscopic glass spheres	0.136	0.139	Microscopic count
Carbon black (Thermax)	7.4	7.69	Nitrogen adsorption[a]

[a] P. H. Emmett and T. De Witt, Ind. Eng. Chem., Anal. Ed. **13**, 28 (1941).

Integration and rearrangement of (2) gives

$$e^{\alpha t}-1=\frac{\alpha}{\beta}\left[\frac{e^{\alpha t}}{(\Delta P)_0}-\frac{1}{\Delta P}\right] \quad (4)$$

where $(\Delta P)_0$ is the initial pressure difference. A plot of $(e^{\alpha t}-1)$ against the quantity in brackets yields a straight line of slope α/β. By definition

$$\beta=\frac{A\epsilon^3}{10V\mu L(1-\epsilon)^2S^2}$$

and hence

$$S=\frac{\epsilon}{1-\epsilon}\left(\frac{A\epsilon}{10V\mu L\beta}\right)^{\frac{1}{2}}. \quad (5)$$

The use of Eq. (1) under varying pressure difference involves an approximation due to neglect of inertia forces; however, the error thus introduced is entirely negligible.

In practice it is most convenient to take two sets of pressure-time readings; one with P varying from 35 to 10 cm Hg and the other in the range of 2 to 0.5 cm Hg using an oil-filled manometer. The latter set of data is used for the $\log\Delta P$ *versus* time plot.

Table I gives two examples of results obtained by the above procedure. Both test plots gave excellent straight lines. The agreement of the surface areas with independent methods of measurement is seen to be entirely satisfactory. The glass microspheres were prepared in the laboratory. A count of 501 particles gave a normal distribution with a mean diameter of 14.21 microns and a standard deviation of 4.60 microns from which the value of S in Table I was calculated.

[1] F. A. Schwertz, J. App. Phys. **20**, 1070 (1949).
[2] P. J. Rigden, J. Soc. Chem. Ind. **66**, 130 (1947).
[3] J. C. Arnell, Can. J. Research **24A**, 103 (1946); **25A**, 191 (1947); **26A**, 29 (1948).

The Pursuit Course

RALPH HOYT BACON
The Perkin-Elmer Corporation, Elmhurst, New York
June 23, 1950

SOME properties of the path of a particle traveling along a curve whose tangent always passes through a point target moving with uniform speed along a straight line have been mentioned by Yuan.[1] Perhaps a fuller description of this curve, called the pursuit course, might be of interest.

Let the path of the target be the y axis, and let the x axis be tangent to the path of the seeker (or pursuer). Then, using Yuan's notation, let

$$\left.\begin{aligned} v_S&=Kv_T\\ dS&=Kdy_T\end{aligned}\right\}, \quad (1)$$

where dS is an element of arc of the pursuit course. For the slope of the tangent of course, at any point, we have

$$\tan\theta=dy_S/dx_S=(y_S-y_T)/x_S.$$

Therefore,

$$\frac{d^2y_S}{dx_S^2}=-\frac{1}{x_S}\frac{dy_T}{dx_S}=-\frac{1}{Kx_S}\frac{dS}{dx_S}=\frac{[1+(dy_S/dx_S)^2]^{\frac{1}{2}}}{Kx_S}.$$

Transposing,

$$\frac{d(dy_S/dx_S)}{[1+(dy_S/dx_S)^2]^{\frac{1}{2}}}=\frac{dx_S}{Kx_S}.$$

Integrating,

$$2\frac{dy_S}{dx_S}=\left(\frac{x_S}{x_0}\right)^{1/K}-\left(\frac{x_S}{x_0}\right)^{-1/K}, \tag{2}$$

where x_0, the constant of integration, is the abscissa of the point of contact of the pursuit curve with the x axis.

Integrating again,

$$\left.\begin{aligned} y_S&=\frac{Kx_S}{2}\left[\frac{1}{1+K}\left(\frac{x_S}{x_0}\right)^{1/K}-\frac{1}{K-1}\left(\frac{x_S}{x_0}\right)^{-1/K}\right]+\frac{Kx_0}{K^2-1} \quad (K\neq 1)\\ y_S&=\frac{x_S^2-x_0^2[1+2\log(x_S/x_0)]}{4x_0} \quad (K=1)\end{aligned}\right\}. \tag{3}$$

For the length of path traveled by the seeker,

$$dS=-\left[1+\left(\frac{dy_S}{dx_0}\right)^2\right]^{\frac{1}{2}}dx_S=-\frac{1}{2}\left[\left(\frac{x_S}{x_0}\right)^{1/K}+\left(\frac{x_S}{x_0}\right)^{-1/K}\right]dx_S.$$

Therefore,

$$\left.\begin{aligned} S&=\frac{K^2x_0}{K^2-1}-\frac{Kx_S}{2}\left[\frac{1}{K+1}\left(\frac{x_S}{x_0}\right)^{1/K}+\frac{1}{K-1}\left(\frac{x_S}{x_0}\right)^{-1/K}\right] \quad (K\neq 1)\\ S&=\frac{x_0^2[1-2\log(x_S/x_0)]-x_S^2}{4x_0} \quad (K=1)\end{aligned}\right\}, \tag{4}$$

where S is measured, plus or minus, from the point of tangency, the point at which $x_S=x_0$.

Combining Eqs. (3) and (4):

$$\left.\begin{aligned} S-y_S&=\frac{K}{K+1}\left[x_0-x_S\left(\frac{x_S}{x_0}\right)^{1/K}\right]\\ &=(x_0^2-x_S^2)/2x_0 \text{ when } K=1\end{aligned}\right\}. \tag{5}$$

Let the distance between the seeker and the target be denoted by P. Then

$$P^2=x_S^2+(y_S-y_T)^2=x_S^2\left[1+\left(\frac{y_S-y_T}{x_S}\right)^2\right]$$

$$=x_S^2\left[1+\left(\frac{dy_S}{dx_S}\right)^2\right]=\frac{x_S^2}{4}\left[\left(\frac{x_S}{x_0}\right)^{1/K}+\left(\frac{x_S}{x_0}\right)^{-1/K}\right]^2.$$

Therefore,

$$P=\frac{x_S}{2}\left[\left(\frac{x_S}{x_0}\right)^{1/K}+\left(\frac{x_S}{x_0}\right)^{-1/K}\right]. \tag{6}$$

Finally, for the radius of curvature of the course,

$$R=\frac{[1+(dy_S/dx_S)^2]^{\frac{1}{2}}}{d^2y_S/dx_S^2}=\frac{Kx_S}{4}\left[\left(\frac{x_S}{x_0}\right)^{1/K}+\left(\frac{x_S}{x_0}\right)^{-1/K}\right]^2. \tag{7}$$

Now, as x_S approaches zero,

$$y_S\rightarrow\frac{Kx_0}{K^2-1}-\frac{Kx_0^{1/K}}{2(K-1)}x_S^{(K-1)/K}, \text{ when } K\neq 1;$$

$$\rightarrow\frac{Kx_0}{K^2-1} \text{ when } K>1;$$

$$\rightarrow\infty \text{ when } K=1 \text{ or when } K<1.$$

$$S\rightarrow\frac{K^2x_0}{K^2-1} \text{ when } K>1;$$

$$\rightarrow\infty \text{ when } K=1 \text{ or when } K<1.$$

$$S-y_S\rightarrow\frac{Kx_0}{K+1}.$$

$$P\rightarrow 0 \text{ when } K>1;$$

$$P\rightarrow\frac{x_0}{2} \text{ when } K=1;$$

$$P\rightarrow\infty \text{ when } K<1.$$

$$R\rightarrow 0 \text{ when } K>2;$$

$$\rightarrow\frac{x_0}{2} \text{ when } K=2;$$

$$\rightarrow\infty \text{ when } K<2.$$

The fact that R approaches zero if K be greater than 2 is well know to every student of the subject.

All of this is an interesting exercise in mathematics; the limitations of its usefulness as a tactical maneuver are pointed by Yuan,[1] who suggests an alternative course.

[1] Luke Chia-Liu Yuan, J. App. Phys. 19, 1122 (1948).

Some Observations on the Back Heating of Magnetron Cathodes

R. T. Young, Jr., L. W. Holmboe, and W. E. Waters, Jr.
National Bureau of Standards, Washington, D. C.
May 24, 1950

In the course of work on *CW* *K*-band magnetrons, we have observed some cathode back heating effects which we consider to be of sufficient interest to be worth recording. The magnetrons were of the interdigital type.[1] The pertinent dimensions of a typical tube are: anode radius—0.16 cm, cathode radius—0.11 cm, active cathode length—0.12 cm, number of teeth—20. The fundamental, first-, second-, and third-order cavity modes for this tube were at 2.9, 1.8, 1.2, and 0.9 cm, respectively. The tubes were designed to operate at 2000–3000 volts, 4000–5000 gauss and 30 ma, some in the second-order mode and some in the first-order mode. A catephoretically deposited thoria on molybdenum cathode was employed.

The experimental results on all tubes were of the same general nature, namely: (a) The tube operated very inefficiently in the desired cavity modes. This type of operation was possible only at low currents, and only under very critical conditions of voltage, magnetic field, and particularly cathode temperature. (b) When either the cathode temperature or the anode voltage was raised, the tube would break discontinuously into an undesired state of oscillation characterized by large back bombardment of the cathode, a marked increase in anode current, and a low power broad-band noise output. The back bombardment power was estimated to be at least 10 watts since it was enough to keep the cathode operating without heater power.

The observed noise output was centered at 1.8 and 1.2 cm, the wave-lengths of the first- and second-order cavity modes, with the stronger signal at 1.8 cm. The frequency separation between the half-power points corresponded roughly to the Q of the cavity. These observations were made on a spectrum analyzer; the power was too small to detect with a thermistor bridge sensitive to 25 microwatts. A signal in the fundamental cavity mode would not be detected because the wave guide was beyond cut-off for the wave-length of that mode. To avoid the frequency limitation of the wave-guide output, a special tube was constructed with a coaxial probe electrostatically coupled to the tooth structure and oriented so as to couple to any degenerate modes. No significant differences in the output could be detected. Our interpretation of these observations is that the cavity acted as a filter for noise generated in the interaction space.

To corroborate the fact that the cavity was not supporting this undesired oscillation, a smooth anode magnetron with identical end space and pole piece configuration, but with a solid cylindrical anode replacing the tooth structure, was tested. The same back bombardment phenomena were observed. A careful search was made for radiation coming out of the cathode leads using a horn feeding *X*-band wave guide, but none could be detected.

Plots of anode voltage *vs.* anode current for the above oscillations were made at magnetic fields from 1200 to 6200 gauss for both the interdigital and smooth anode tubes. They were all similar to the *V-I* curves of a normal cavity magnetron. The curves of voltage *vs.* magnetic field also were similar to the *V-B* curves of a normal cavity magnetron. That is, they were straight lines, approximately tangent to the cut-off parabola. These results held for tubes with a variety of end space configurations, magnetic field distributions, and ratios of anode-cathode radii.

It would appear that a state of oscillation is involved which is primarily a function of the space charge conditions in the interaction space. Since we have not identified any frequency associated with the oscillation, it is difficult to draw any conclusions as to its nature. One supposition is that there is a generation of noise alone. However, one cannot rule out the possibility that the observed signals are noise side bands of a coherent signal. From the linearity between V and B, one concludes that the condition for the oscillation depends upon the angular velocity of the electrons.

The phenomenon of back heating in smooth anode magnetrons has been noted by several authors[2-4] and has generally been attributed to some type of oscillation. There is no completely satisfying explanation available. Using a rising sun design, Columbia Radiation Laboratory[5] has obtained up to 30 watts from K-band magnetrons under roughly the same operating conditions as ours. Their reports briefly mention some back heating effects which probably are due to the same type of oscillation as reported here, but these effects do not prevent normal oscillation in their tubes.

It appears that in the magnetrons we worked with this undesired oscillation is sufficiently pronounced to prevent normal operation of the tube. An understanding of this type of oscillation may be of fundamental importance in the design of low voltage, short wave-length magnetrons.

[1] F. H. Crawford and M. D. Hare, Proc. I.R.E. **35**, 4, 361–369 (1947).
[2] E. G. Linder, Proc. I.R.E. **24**, 633–653 (April, 1936).
[3] E. G. Linder, Proc. I.R.E. **26**, 346–371 (March, 1938).
[4] E. C. S. Megaw, Nature **132**, 854 (December, 1933).
[5] Columbia Radiation Laboratory Quarterly Progress Report, December 31, 1947 and 1948 reports, Signal Corps Contract W-36-039 sc-32003.

Some Magneto-Acoustic Effects in Nickel*

T. F. ROGERS
Air Force Cambridge Research Laboratories, Cambridge, Massachusetts
AND
S. J. JOHNSON
Crystal Research Laboratories, Inc., Hartford, Connecticut
June 1, 1950

AN investigation has been made of effects produced by a magnetic field upon the ultrasonic propagation properties of certain ferromagnetic substances, of which nickel is reported upon here. A rod of commercially pure—99.4 percent—"A" nickel 45.8 cm long and 1.25 cm in diameter employing a.c.-cut quartz crystal transmitting and receiving transducers, exhibits an "electrical length" of some 170 μsec. when acting as a propagation medium for a short-pulsed, 1-Mc sonic carrier frequency, with the transmitting crystal generating a transverse (shear) wave within the rod. After having been heated to approximately 1600°F, a red heat assuring demagnetization, and allowed to cool slowly to room temperature, 70°F, the rod was subjected to a longitudinal magnetic field throughout its entire length; changes in its physical length were measured directly, while transmission (delay) time and relative attenuation were measured by means of an oscilloscope reproduction of the received pulse, both as functions of the applied field strength.

Figure 1 indicates the change—a decrease—in transverse wave propagation time delay (electrical length) ΔT, relative to the original time delay T, as a function of the field intensity in oersteds. The observed changes in delay of approximately 10^3 times that expected from the magnetostrictive (Joule) effect in nickel can, apparently, only be explained as signifying a marked increase in the acoustic velocity of propagation within the rod while it is under the magnetic field stress.

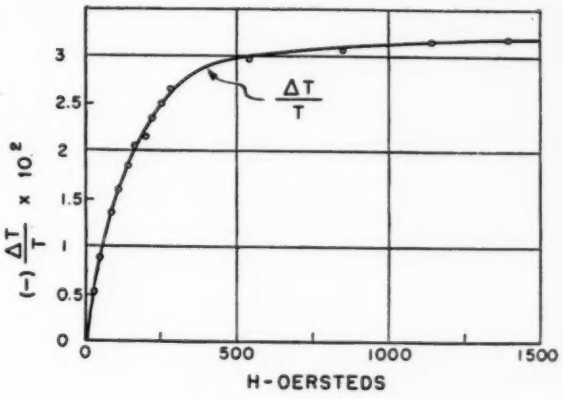


FIG. 1. Some magneto-acoustic effects in nickel.

The velocity of propagation of a transverse vibration is given as:[1]

$$V_t=\left[\frac{\mu}{\rho}\right]^{\frac{1}{2}}=\left[\frac{E}{2\rho(1+\sigma)}\right]^{\frac{1}{2}},$$

where μ is the modulus of rigidity, E is Young's modulus, ρ is the density, σ is Poisson's ratio, and $E=2\mu(1+\sigma)$. With the applied field, the change in density is considered negligible as a factor contributing to the velocity change, since it amounts to a factor of less than 10^{-4}, and, therefore, the increase in velocity must be attributed to a change in the value of the modulus of rigidity; the work of various groups[2] on the variation of E with magnetization (the so-called ΔE effect) would also appear to warrant this conclusion.

A large decrease in attenuation was also noted upon application of the field, with the received pulse increasing in amplitude some 10^2 times its normal value. Figure 2 presents a representative

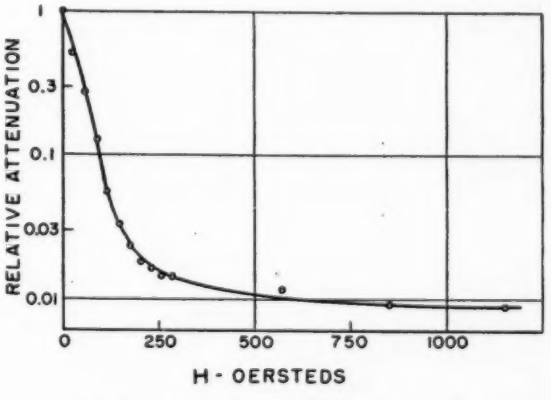


FIG. 2. Some magneto-acoustic effects in nickel.

change in attenuation throughout the transmission system as a function of H with the output pulse taken as unity height for $H=0$. It has been found that, for certain metallurgical conditions of the sample, irregularities are noted in the attenuation *vs.* H curve, with treatment of the rod immediately previous to test influencing the general march of values to a considerable extent. While possibly connected with a reorientation of the nickel crystalline structure, at this time, especially in view of the fact that ultrasonic attenuation effects in solid media are not altogether understood,[3] no adequate explanation of this decrease in attenuation presents itself.

It is planned that work on nickel, as well as other magnetic materials, will be continued and extended. These initial results are presented at this time for reasons of the natural interest in

such originally noted phenomena and in the expectation that this investigative technique will lend itself to a study of acoustic propagation in magnetic media and the structure of the media themselves.

* This work was supported by the Air Force Cambridge Research Laboratories, Cambridge, Massachusetts.

[1] L. Bergmann, *Ultrasonics* (John Wiley and Sons, Inc., New York, 1939), p. 162.

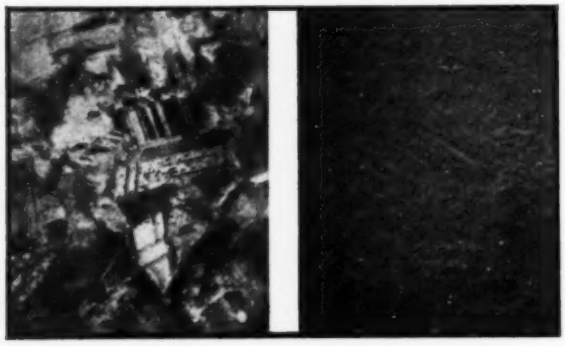
[2] E. Giebe and E. Blechschmidt, Ann. d. Physik LPZ (v) **11**, 905 (1931); S. Siegal and S. L. Quimby, Phys. Rev. **49**, 663 (1936); R. Kimura, Proc. Phys. Math. Soc. Japan **21**, 786 (1939).

[3] C. Kittel, *Reports on Progress in Physics* (The Physical Society, London, 1948), Vol. XI (1946–47), p. 205.

Electron-Microradiography of Electrodeposited Metals

ROLF WEIL AND HAROLD J. READ
Division of Metallurgy, Pennsylvania State College, State College, Pennsylvania
August 2, 1950

IN the course of research on the structure of electrodeposited metals, a new method of examining specimens directly in the electron microscope, without the aid of replicas, was developed. This method is most valuable in the study of the initiation of plating, revealing some features not seen when conventional techniques are employed. As annealed electroplated metals are essentially the same as annealed metals prepared from the melt, the method described here should also be a useful tool in the study of other metallurgical phenomena.



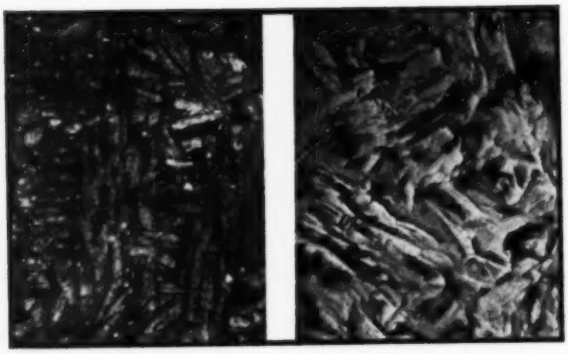


FIG. 1. (Upper left.) Electron-microradiography of electrodeposited nickel on copper. 5000×.

FIG. 2. (Upper right.) Electron-micrograph from collodion replica of specimen shown in Fig. 1. 5000×.

FIG. 3. (Lower left.) Electron-microradiograph of electrodeposited nickel on zinc. 5000×.

FIG. 4. (Lower right.) Electron-micrograph from collodion replica of specimen shown in Fig. 3. 5000×.

To date, nickel films plated on copper and zinc have been examined. The method of preparing the samples was as follows: The starting cathode was cleaned by the usual procedures for electroplating, and a deposit approximately 5×10^{-4} cm thick laid down. A piece the size of an electron-microscope screen was cut from the plated cathode, and placed in a suitable solvent to dissolve the basis-metal. For nickel-plated copper, a solution of 500 g/l CrO_3 and 50 g/l H_2SO_4 dissolved the copper, leaving the nickel. A solution of five percent by volume of concentrated HCl in water served for dissolving zinc basis-metals. When the original cathode was dissolved, the nickel film was washed and placed on a screen, ready for examination.

The accompanying photographs (Figs. 1–4) show the structures of nickel films deposited on copper and zinc respectively as seen by electron-microradiography and by the collodion replica techniques. The films were deposited under the same plating conditions, thus illustrating the effect of the basis-metal. The direct method resolved the structure far more distinctly than collodion replicas.

For convenience in referring to pictures made by the above technique, we should like to suggest the term "electron-microradiography" so as to distinguish conveniently such pictures from those obtained by the replica technique.

The Limiting Hydrostatic Tension of Water Near 0°C

J. C. FISHER
General Electric Research Laboratory, Schenectady, New York
August 3, 1950

RECENTLY L. J. Briggs has determined the limiting negative pressure in water as a function of temperature.[1] Above 5°C his values are in the neighborhood of 270 negative bars, about 20 percent of the theoretical value calculated from nucleation theory.[2] Below 5°C the limiting tension was observed to drop rapidly with decreasing temperature to a value near zero just above 0°C.

The rapid decrease of limiting tension can be explained in terms of the nucleation of ice under reduced pressure. Assume that ice nuclei form readily and grow when the solid phase is stable, and that ice nucleates the vapor phase. The assumption that ice nucleates water vapor is consistent with the lower solubility of gasses in ice and the tendency of gas bubbles to form at the surface of a growing crystal. Then near 0°C the limiting tension that water will withstand is that required to raise the freezing point to the testing temperature.

A negative pressure of 300 bars raises the freezing point of water 2.7°C, in fair agreement with Briggs' observations. A possible check of the ice nucleation hypothesis would be to determine the supercooling limit of the water at atmospheric pressure. If appreciable supercooling is observed, it is unlikely that ice nuclei form as readily as would be required to account for the decrease in limiting tension.

[1] L. J. Briggs, J. App. Phys. **21**, 721 (1950).
[2] J. C. Fisher, J. App. Phys. **19**, 1062 (1948).

Local Order in Solid Cu_3Au

R. A. ORIANI
Research Laboratory, General Electric Company, Schenectady, New York
June 5, 1950

RECENT x-ray data by Cowley[1] on Cu_3Au at temperatures above the critical temperature of superlattice formation enable the degree of local order computed from x-ray measurements to be compared with that calculated from thermodynamic

activity coefficients measured by Weibke and Quadt.[2] Such a comparison is of great interest, for this is the first system, to the writer's knowledge, for which two independent sets of data at overlapping temperatures have been obtained from which local order may be computed by two different methods of analysis. This note briefly summarizes calculations of local order at the composition Cu_3Au at 500°C, over 100° higher than the critical temperature of superlattice formation.

1. *X-ray data.*—Cowley's short-range order parameters for the first coordination shell at 405, 460, and 550°C were employed to interpolate the value of $\alpha=-0.142$ at 500°C, since the e.m.f. data for 500°C were used. In Cowley's terminology,

$$\alpha=1-(P_{l,m,n}/m_{Cu}), \tag{1}$$

where $P_{l,m,n}$=probability that the atom with coordinates l, m, n with respect to an Au atom should be a Cu atom; m_{Cu}=proportion of Cu atoms=0.75 for Cu_3Au. For complete randomness, $\alpha=0$ so that $P=m_{Cu}$. For perfect order, $\alpha=1$ for even-numbered shells, and $\alpha=-\frac{1}{3}$ for odd-numbered shells. The experimental value of α, and hence the degree of local order, depends only on the precision of the experimental measurements and on the adequacy of the three-dimensional Fourier analysis of the diffuse background scattering of x-rays. From $\alpha=-0.142$, one easily computes that (a) for Cu at the origin, the average number of Cu atoms in the first coordination shell is 8.55, and of Au atoms, 3.45; (b) for Au at the origin, the average number of Cu atoms in the first coordination shell is 10.3, and of Au atoms, 1.7. Hence, x=the ratio of the number of Au−Au bonds to the total number of bonds=0.0354, whereas for random distribution $x=0.0625$. Hence, there are more bonds between dissimilar atoms than the number expected from a random distribution of atoms; the percent reduction of Au−Au bonds from that characteristic of a fully disordered structure is 43.4 percent.

2. *Activity data:* A. *First method.*—From Weibke and Quadt's 500°C e.m.f. data, a smoothed curve of $\ln\gamma_{Cu}$ *vs.* mole fraction of Cu was constructed; the activity coefficients, γ, are referred to pure copper as the standard state. From this curve $\gamma_{Cu}=0.79$ at 0.75 mole fraction of Cu was obtained. The activity coefficient for gold was calculated as 0.143 by graphical integration of the Duhem-Margules equation.

Takagi[3] has developed a relation between x, the fraction of Au−Au bonds, and ΔF°, the free energy change for the bond reaction

$$\text{Au}-\text{Au}+\text{Cu}-\text{Cu}\rightarrow 2(\text{Au}-\text{Cu}). \tag{2}$$

Birchenall[4,5] independently derived the special case applicable to Cu_3Au, i.e.,

$$\frac{(1-4x)^2}{2x(1+2x)}=4e^{-\Delta F^\circ/RT}. \tag{3}$$

In order to compute ΔF°, we make use of regular solution theory,[6] relating the activity coefficient of A to the difference between the partial molar internal energy of vaporization of an ideal solution of A in B and that of the pure metal. Following Birchenall,[4] we neglect the volume change in going from ideal to regular solution, and assume as a first approximation that the atoms in the solution are distributed randomly in order to compute the partial molar internal energy of vaporization of the solution. Then,

$$\Delta F^\circ=\Delta\bar{F}_{Cu}+\Delta\bar{F}_{Au}\cong\Delta\bar{E}_{Cu}+\Delta\bar{E}_{Au}=RT(\ln\gamma_{Cu}+\ln\gamma_{Au}),$$

and by use of (3), a value of x is obtained. The whole calculation may be repeated, correcting for the randomness assumed by using this value of x; the second approximation yields ΔF° for reaction (2)=−820 cal./mole of bonds, and x=fraction of Au−Au bonds =0.0453, representing a 27.5 percent reduction in the number of Au−Au bonds from the number characteristic of randomness. In order to evaluate the kind of variation introduced by an error in the experimental data, the first part of this calculation was repeated using a γ_{Cu} some 3 percent lower than the 0.79 employed above, since this is about the scatter of the data. The value of x thus computed is less than 2 percent larger than that obtained above.

B. *Second method.*—Lo-Ching Chang,[7] using Takagi's development has shown that if there is no volume change on mixing, and if the pairwise energies of the atoms are independent of composition, the heat of solution,

$$\Delta H=(x_{11}-m_1)(v_{11}+v_{22}-2v_{12}), \tag{4}$$

where x_{11}=fraction of 1−1 bonds, m_1=mole fraction of component 1, and v=energy of a pair of atoms. Transcribing this into our nomenclature, we have

$$\Delta H=(1/2)Z\Delta F^\circ(0.25-x), \tag{5}$$

where Z=coordination number=12. This method of treatment is very similar to that of part (A), since Eq. (5) also makes use of the regular solution approximation. If we use the value of $\Delta H=-1000$ cal./mole obtained from Weibke and Quadt's activity data, and solve (3) and (5) simultaneously, we obtain

$\Delta F^\circ=-816$ cal./mole for reaction (2)
$x=0.0455$
27.2 percent reduction in the number of Au−Au bonds.

In view of the difficulties of the measurements and the approximations involved in the thermodynamic treatment, the agreement among the three methods is quite satisfactory, and lends confidence to the use of activity coefficients in this manner. A similar satisfactory comparison between x-ray and thermodynamic measurements of local order on the Ag−Au system, though at two different temperatures, has been made by B. L. Averbach reported by C. E. Birchenall.[8]

[1] J. M. Cowley, J. App. Phys. **21**, 24 (1950).
[2] F. Weibke and U. v. Quadt, Zeits. f. Elektrochemie **45**, 715 (1935).
[3] Y. Takagi, Physico-Math. Soc. Japan **23**, No. 1, 44 (1941).
[4] C. E. Birchenall, A.I.M.E. Tech. Pub. No. 2169 (1947).
[5] C. E. Birchenall and C. H. Cheng, Metals Trans. **185**, 428 (1949).
[6] J. H. Hildebrand, *The Solubility of Non-Electrolytes* (Reinhold Publishing Corporation, New York, 1936).
[7] Lo-Ching Chang (unpublished).
[8] C. E. Birchenall, Am. Soc. Metals, Cleveland, 159 (1950).

Formation of Annealing Twins during Grain Growth

R. L. FULLMAN
Research Laboratory, General Electric Company, Schenectady, New York

June 5, 1950

A DOMINANT FEATURE of the microstructures of most annealed metals and alloys with a face-centered cubic crystal structure is the presence of many straight-sided crystals that have a twinned orientation relative to the grains in which they appear. Mathewson[1] proposed that these "annealing twins" form during recrystallization, by lateral growth of "twin faults" produced by the slip process. Carpenter and Tamura[2] found that annealing twins may form under conditions where it appears doubtful that slip could be of importance. They suggested that annealing twins form by accidents of growth when grain growth occurs. Burke[3] found that, during grain growth, annealing twins appear, one boundary at a time, in the corners of growing grains. During continued grain growth the distance between twin boundaries was not observed to change.

The interfacial free energy of the boundary between adjacent grains is a function of the orientation relationship between the grains.[4] During grain growth, changes occur in the identity of a grain's neighbors, and the interfacial free energies of the grain boundaries also change. It is suggested that an annealing twin forms in the corner of a growing grain when the orientation relationships between the twin and the grain's neighbors lead to a smaller total interfacial free energy than do the orientation rela-

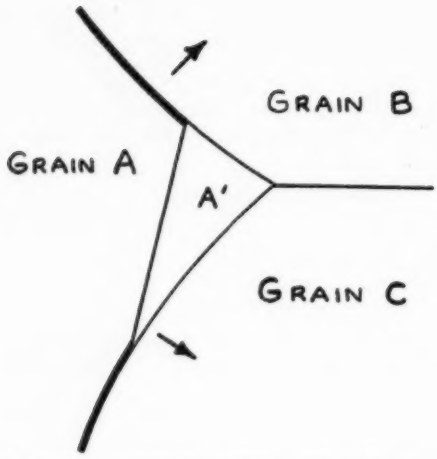


FIG. 1. Schematic representation of the changes in grain boundary free energy associated with the formation of a twin, A', in an infinitesimal region in the corner of a growing grain, A. Grain boundaries with high interfacial free energies are represented by wide lines.

tionships between the grain itself and its neighbors.* A twin boundary then forms in a corner of the grain, and the grain continues to grow as a twin of the original crystal. This could occur at various times during the growth of a grain, when the identity of one or more of the neighboring grains changes. The conditions required for a decrease in free energy may be clarified by reference to Fig. 1, which represents schematically a small region where atoms are being added to the corner of a growing grain, A, and removed from grains B and C. The small arrows indicate the directions of grain boundary migration, and the wide lines indicate boundaries with relatively high interfacial free energy. The free energy will be decreased by formation of a twin, A', in the corner of grain A, if

$$l_{A'B}\sigma_{A'B}+l_{A'C}\sigma_{A'C}+l_{AA'}\sigma_{AA'}<l_{A'B}\sigma_{AB}+l_{A'C}\sigma_{AC}, \quad (1)$$

where l and σ represent the areas and interfacial free energies, respectively, of the boundaries indicated by the subscripts. Annealing twins would form in grain corners during grain growth when the geometry and orientation relationships are such that inequality l is satisfied.

During grain growth the identity of a grain's neighbors may change by several mechanisms. Such changes may lead to the formation of new twins or the removal of existing twins. Details of the mechanism of formation and disappearance of annealing twins, and experimental evidence supporting the theory, will be given in forthcoming papers. Experiments designed to detect the difference in the interfacial free energies of grain boundary segments on opposite sides of twin boundaries newly formed in grain corners indicate that the free energy actually is smaller toward the grain corners. The dependence of the abundance of annealing twins in copper on the preferred orientation and grain size was found to be in agreement with predictions made on the basis of the proposed theory of twin formation and disappearance. The ratio of the interfacial free energy of twin boundaries to that of grain boundaries has been measured in copper and in aluminum. The values found were 0.036 ± 0.006 and 0.21 ± 0.05, respectively. Annealing twins rarely are observed in aluminum. The difference observed in the abundance of annealing twins in copper and aluminum is in accordance with the proposed theory and the difference found in the twin boundary to grain boundary interfacial free energy ratios.

This letter is an abstract of a dissertation presented to the Yale School of Engineering in partial fulfillment of the requirements for the degree of Doctor of Engineering. The investigation was supported by the United States AEC under Contract No. W-31-109-ENG-52.

[1] C. H. Mathewson, Trans. A.S.M. **32**, 38 (1944).
[2] H. C. H. Carpenter and S. Tamura, Proc. Roy. Soc. **113A**, 161 (1926).
[3] J. E. Burke (to be published).
[4] C. G. Dunn and F. Lionetti, Trans. A.I.M.E. **185**, 125 (1949).

* This concept was suggested by J. C. Fisher, who has also assisted in the further development of the theory.